nature

www.nature.com/nature

Vol 446 | Issue no. 7136 | 5 April 2007

Museums need two cultures

The resignation of the head of the Smithsonian Institution highlights a misguided tendency for museums to focus on communication at the expense of research. It also offers the chance of a fresh start.

Corporate managers like mission statements; scientists generally don't. Academic freedom often sits uneasily alongside the goal-driven culture of the private sector. Researchers would prefer not to be told what to do. It's interesting, therefore, that scientists at the Smithsonian Institution are so proud of their mission statement — especially as they have only just welcomed the resignation of Larry Small, a leader who angered many by trying to impose a corporate ethos on the organization (see page 594).

But the vision they share is not that espoused by the departing Small. They are proud instead to recite the mission statement of the British scientist James Smithson, who funded the creation of their institution in 1846 for the "increase and diffusion of knowledge". Small seemed either to not understand, or to not care too much, about the fact that the museum should remain committed to both public outreach and to new research. Worryingly, the leaders of some other prominent museums also have this fault.

The scientists' desire to fulfil both strands of Smithson's vision is notable. Many researchers in purely academic institutions often shy away from outreach work. Some lack the skills. Others fear, often rightly, that funders will not reward them for time spent communicating what they do. Museums such as the Smithsonian are among the only places where scientists are required to do both. They are historical oddities, as it is now extremely rare for such dual-purpose organizations to be created. Yet the combination of missions leads to a special form of science communication.

When outside professionals are drafted in to build exhibits, as is the trend in many natural-history and science museums, there are benefits. But there are also risks unless professional researchers continue to play a leading role. Be they palaeobiologists or historians of science, their involvement should lead to the creation of exhibits that intimately reflect the science behind the display in a way that other forms of science communication, such as science writing, cannot. Some of these displays are now badly dated, but the galleries of dusty fossils that gave scientific curators a bad name were dumped years ago by the more innovative institutions.

The Natural History Museum in London, for example, is creating a centre that will allow visitors to see parts of the museum's extraordinary research collections and to interact with the scientists who work on them. At the Smithsonian, researchers want to channel real-time data from animal monitoring experiments at the institution's field stations into exhibits about the species involved.

Such projects happen naturally in organizations that do both science and science communication, yet the research side of many of these institutions is suffering. Small paid too little heed to it during his seven years at the Smithsonian. The collections division at the Science Museum in London, which houses curators who also do historical research, has been gutted over the past 20 years. In others, such as the Academy of Natural Sciences in Philadelphia, which is profiled on page 605, parts of the collection have been sold to make up for funding shortfalls.

"Researchers can create exhibits that reflect the science in ways that other forms of science communication cannot."

This gloomy trend now needs to be reversed, and there are signs of hope. For example, the new head of the Philadelphia academy, ornithologist William Brown, has a track record of valuing and understanding science. As the Smithsonian's regents begin their search for a successor to Small, they should look at Brown's plans, or those of London's Natural History Museum, because museums that combine science and outreach are vibrant and unique places. Exhibits get the media attention and attract donors. But take away the in-house scholars who help build them, and museums take another step towards becoming little more than theme parks. ■

Timber and tapirs

A biodiversity conservation project needs support, a watchful eye, and maybe even a long-snouted ally.

The Malaysian state of Sarawak, on the island of Borneo, is pushing forward with a unique conservation project. On a piece of land nearly twice the size of Luxembourg, it is going to juxtapose three strikingly different uses for land: plantations of acacia, a pulp-producing, non-native monoculture tree, for logging; a natural forest for local indigenous groups to cut and farm; and a biodiversity conservation zone. The biologists in charge of the conservation zone are optimistically thinking that even the plantation zone, if designed and logged systematically, could play a role in supporting species (see page 608).

For many long-time observers of conservation issues, the initial response to such a project may be to roll their eyes and think: "Here we go again." There have, after all, been many timber companies — particularly ones operating in Indonesia, which takes up most of the rest of Borneo — that claimed to be logging 'sustainably' when in fact they were destroying natural forests.

Such scepticism is understandable, but it doesn't make the Sarawak project unworthy of support. Indeed its design gives some grounds for optimism. The conservation biologists involved in it have some economic leverage, as timber companies are subject to increasing pressure from buyers to supply timber that is grown and logged in ways that minimize environmental impact. This was probably the

government's central motivation in 'going green' with the planted forest project in the first place.

Even so, the conservation project's organizers have a tough job. If the loggers or timber companies do not seem to be playing by the rules — for example, failing to protect conservation zones as promised, or not using low-impact logging techniques — they need to cry foul to the media or the appropriate government agency. This is the only way to ensure that the project pays more than lip service to its environmental goals. But given the delicate balance of conservation politics and economic goals in Sarawak, it will be difficult to do.

There are other ways to help keep the project on the right path. The conservation aspect depends on proving that there are species worth saving and finding ways to protect them. Ecologists and taxonomists from around the world can help by studying habitat, taking specimens and reporting back. Rather than popping in just to get samples for their collections, they should commit to periodic visits so that changes in both the natural forest and the plantation can be monitored systematically over time in different seasons. Many researchers have already started doing this, but more participation is needed to document the flora and fauna at stake.

One of the most creative proposals to make the most of the experiment in Sarawak concerns the tapir, a large-bodied, short-legged herbivore with splayed feet and a prehensile snout. There is archaeological evidence that Malayan tapirs (*Tapirus indicus*) were present on the island as recently as 60 years ago, and some researchers have argued that the conservation zone would be a perfect opportunity to bring them back. The mammal's significance in ecological terms might be small, but its restoration to its natural habitat could have considerable symbolic impact. Such reintroductions of large mammals are rare. Globally the Malayan tapir alternates between vulnerable and endangered on conservation lists. Its disappearance from Borneo was probably due to a loss of habitat following uncontrolled logging. What better way to show that logging is now controlled and its habitat protected?

Sadly, Sarawak's biodiversity programme will probably not be able to provide habitat for many other species that need it. Some will be lost. This is an inevitable outcome of a simple reality: to governments such as Sarawak's, as well as to the people who live in and around the forests, the use of the forest is a crucial part of their livelihoods.

The current project is ambitious and, as its organizers will readily admit, success is by no means assured. They haven't even agreed how success should be measured. But subject to these caveats, the project holds considerable promise as a model for biodiversity conservation in a southeast-Asian timber industry that badly needs one. ■

South Africa's mentors

Announcing this year's *Nature* awards for scientific mentoring.

Nowhere is strong mentoring for young scientists more urgently required than in South Africa. The Centre for Research on Science and Technology at Stellenbosch University has shown that the country's research output declined in nominal terms between 1987 and 2001, and almost halved as a proportion of global scientific output, as measured by the production of articles listed by Thomson Scientific (ISI). South Africa's scientific workforce is ageing, with few young researchers being attracted into the system. In 1990, only 20% of peer-reviewed articles published in South Africa were written by people older than 50, but this figure had risen to 49% by the year 2000.

Additionally, there was no significant change in the demographic profile of publishing scientists between 1990 and 2000, with white scientists continuing to produce more than 90% of peer-reviewed articles. Women, who make up 40% of the academic staff at the country's universities, are responsible for authoring only 17% of articles. The National Research Foundation has attempted to address these problems by providing additional sources of funding to support black and female faculty members, and for black postgraduate students.

Originally the legacy of apartheid-based education, the dearth of black researchers has tragically continued into South Africa's democratic era. The post-apartheid government has been woefully unsuccessful in improving the maths and science education of black Africans since it took office in 1994: a 2004 report by the Johannesburg-based Centre for Development and Enterprise revealed that the same number of black school-leavers achieved maths grades sufficient for university entrance in science in 2002 as in 1991.

The country's universities are experiencing problems at two levels. First, far too few students — and particularly black students — enrol for undergraduate degrees in science, medicine and engineering. Second, relatively few students who complete first degrees continue their studies to doctoral level.

Whereas the former problem will require a long-term solution, the latter one could be solved relatively easily by substantially increasing the value of postgraduate bursaries. This is a measure the Department of Science and Technology is going to have to take if it is serious about its stated aim of increasing the number of doctoral graduates — currently just over 1,100 annually — by an order of magnitude.

But numbers are not the whole story. Equally important for South Africa's future is the quality of science produced, and a critical factor here is the quality of mentoring that fledgling researchers receive. For all its problems, South Africa's universities include some with a strong scientific culture. There is no reason to suppose that, as a proportion of its academic population, there are fewer great science mentors in South Africa than anywhere else.

At *Nature*, we want to highlight them and the good things they do. Competitions for *Nature*'s mentoring awards, initiated in 2005 and so far held in the United Kingdom and Australasia, are intended not only to celebrate the winners, but also to show the way for others. So we urge readers who know of outstanding mentors in South Africa to ensure that they are nominated. Two awards of 60,000 rand (US$8,300) will be given: one for lifetime achievement and one for a track record in mid-career.

The closing date for nominations is 31 May. Details of the competition, the judges and how to nominate candidates for the awards can be found at www.nature.com/nature/mentoringawards/southafrica. ■

RESEARCH HIGHLIGHTS

Eroding principles

Geology **35,** 303–306 (2007)

A survey in southern Ecuador quantifies the dramatic effects of removing surface vegetation on soil erosion in tropical mountain regions. But the good news is that restoring dense plant cover, even if it is not native vegetation, cuts erosion back to near its natural rate.

Veerle Vanacker, now at the Catholic University of Louvain, Belgium, and her colleagues studied isotopes in river sediments to show that the natural erosion rate, averaged over the past 20,000 years, has been about 150 tonnes per square kilometre per year. The team showed that on slopes cleared by humans this rate increases up to 100-fold. They found reduced erosion rates for land covered by pine or eucalyptus plantations.

V. VANACKER

GENETICS

Birds do it differently

J. Biol. **6,** 2 (2007)

Gender parity may not be as important in the bird world as it is for other species, report Arthur Arnold of the University of California, Los Angeles, and his colleagues.

Animals such as humans and fruitflies have evolved various types of 'dosage compensation' to ensure that males and females express genes on the sex-related X chromosome(s) at the same level, even though females have two copies of the X to males' one.

Dosage compensation is thought to be crucial for the survival of a species. But Arnold's team found that zebra finches and chickens defy this expectation. On average, the males of these bird species express sex-linked genes at higher levels than females (in birds, it's the males that have two identical sex chromosomes).

The authors speculate that the unequal expression could be a novel way to generate sexual dimorphism.

INFECTIOUS DISEASES

HIV modelled in mice

J. Exp. Med. doi:10.1084/jem.20062411 (2007)

Scientists in the United States report the development of a mouse model of HIV that, they say, could facilitate the search for new treatments and preventative measures against the disease.

It has been difficult to find appropriate small-animal models of HIV infection — especially those that model rectal transmission. J. Victor Garcia at the University of Texas Southwestern Medical Center in Dallas and his colleagues approached this problem by transplanting human tissue — including bone-marrow stem cells — into mice. The resulting mice developed many human cells, including the immune system's T cells. Garcia's team now reports that the mice respond to rectal HIV infection in a similar fashion to humans, including rapid depletion of gut T cells and the production of antibodies against the virus.

MATERIALS SCIENCE

Alphabet soup

J. Phys. Chem. Lett. C **111,** 4477–4480 (2007)

Crisply rendered particles shaped like letters of the alphabet float about in the soup (pictured below) prepared by Thomas G. Mason and Carlos Hernandez at the University of California, Los Angeles.

The soup shows off a lithography method

AM. CHEM. SOC.

developed by the team, which can mass-produce tiny particles with intricate shapes. The particles are formed in a layer of polymer on a silicon wafer, then float freely when an underlying layer is dissolved. Each of the letters measures about 4 by 7 micrometres — small enough that the particles dance the thermally induced random waltz of brownian motion when dispersed in a fluid.

The method could make any number of shapes, providing a tool to study how differently shaped particles interact, assemble, reconfigure and jam together.

ASTRONOMY

Star bright

Astrophys. J. **659,** L13–L16 and *Astrophys. J.* (submitted) preprint at http://arxiv.org/abs/astro-ph/0612617 (2007)

Astronomers are puzzling over what is probably the brightest supernova ever observed — a stellar explosion that, at its peak, pumped out almost 100 billion times more energy per second than is emitted by our Sun.

Telescopes on the ground and in space have tracked the fate of the supernova, SN 2006gy, since it was first detected in September 2006. Having analysed these data, two teams now present differing theories for what might have happened.

Eran Ofek of the California Institute of Technology in Pasadena and his colleagues suggest that the event's unusual brightness was due to the exploding star being enclosed in a dense gas cloud, which converted the mechanical energy of the burst into radiation. In a second paper submitted to the same journal, Nathan Smith of the University

of California, Berkeley, and his colleagues argue that it was possibly a 'pair instability' supernova — the explosion of an extremely massive star, of a type that theorists might have expected to collapse quietly to form a black hole.

BIOTECHNOLOGY

Blood for all

Nature Biotech. doi:10.1038/nbt1298 (2007)

Bacterial enzymes that efficiently convert blood from groups A, B and AB into the 'universal' O group may lead to safer blood transfusions.

Red blood cells from groups A, B and AB contain antigens that cause life-threatening reactions if transfused into people with a different blood group. Henrik Clausen of the University of Copenhagen, Denmark, Gerlind Sulzenbacher of the Universities of Aix-Marseille, France, and their colleagues have identified two novel glycosidase enzymes that strip away these antigens. The enzyme-treated blood could then be used like group-O blood — group-O cells do not carry A or B antigens, so they can be safely given to anyone. The technology now needs to be tested in clinical trials.

MATERIALS SCIENCE

Waves of honey

Langmuir **23,** 3732–3736 (2007)

Even the simplest ingredients can give rise to complex patterns. Mark Buchanan of the University of Oslo, Norway, and his team have tackled one family of patterns — those formed when fluids containing suspended particles flow down a vertical surface.

Yoghurt running down the side of a pot, for example, is cut through by vertical, branched channels, whereas honey tends to break up into wavy horizontal bands. The researchers show that particle size determines which type of pattern is produced.

Smaller particles leave behind vertical channels, because the downward flow of the film becomes focused along certain paths by a feedback mechanism that amplifies flow rate. Wavy horizontal bands appear in suspensions of larger particles, as particles get trapped by random imperfections on the vertical surface and jam together to form stress-bearing arches. The pictures above show these effects; the fraction of large particles increases from left to right.

AM. CHEM. SOC.

MOLECULAR ELECTRONICS

A promising line-up

Nature Nanotechnol. doi:10.1038/nnano.2007.77 (2007)

Arrays of thousands of perfectly linear and parallel carbon nanotubes have been turned into transistors by John Rogers of the University of Illinois in Urbana-Champaign and his co-workers.

The researchers grew their nanotubes on quartz decorated with iron-oxide stripes that catalyse the tubes' growth. The tubes line up along the crystal axis of the substrate. Each tube could be divided into many transistors by careful placement of metal electrodes on the array's surface. Devices built in this way showed good performance — particularly in having high 'mobility', a measure of how readily current can move through them. The nanotube arrays could also be transferred from the quartz onto plastic substrates, which could be useful in flexible displays.

CELL BIOLOGY

Predictions on target

Nature Struct. Mol. Biol. doi:10.1038/nsmb1226 (2007)

A novel model might help biologists to predict the targets of microRNA molecules in animals. This is notoriously difficult, because in animals microRNAs can block the translation of messenger RNA (mRNA) into proteins without matching precisely the sequence to which they bind.

Devised by Ye Ding at the Wadsworth Center of the New York State Department of Health and Victor Ambros of Dartmouth Medical School in Hanover, New Hampshire, and their colleagues, the model takes into account the structure of the target mRNA, calculated with a program called Sfold. It assumes that microRNAs first find a short open stretch of mRNA to bind to, then unfurl the surrounding mRNA to complete binding.

This approach predicted strong interactions for some known microRNAs and their targets. It also correctly ruled out 11 of 12 targets predicted for a particular microRNA by sequence matching that were not borne out by experiments.

JOURNAL CLUB

Adina Paytan
Stanford University, California

A palaeoceanographer worries not about corals, but about coral reefs.

To understand what the consequences of human-induced CO_2 increases might be, I study how atmospheric CO_2 concentrations fluctuated in the past.

One outcome of high atmospheric CO_2 that is inevitable is ocean acidification. Atmospheric CO_2 dissolves in sea water, lowering the pH of the ocean's surface layer.

We expect this to create problems for marine creatures that precipitate their skeletons from calcium carbonate, because the mineral dissolves in acid. Some researchers have suggested that scleractinian corals might even be driven to extinction.

But what does the geological record tell us? Corals' reef-building fossils have appeared and disappeared over the past 200 million years and despite periods of elevated atmospheric CO_2, the organisms did not go extinct.

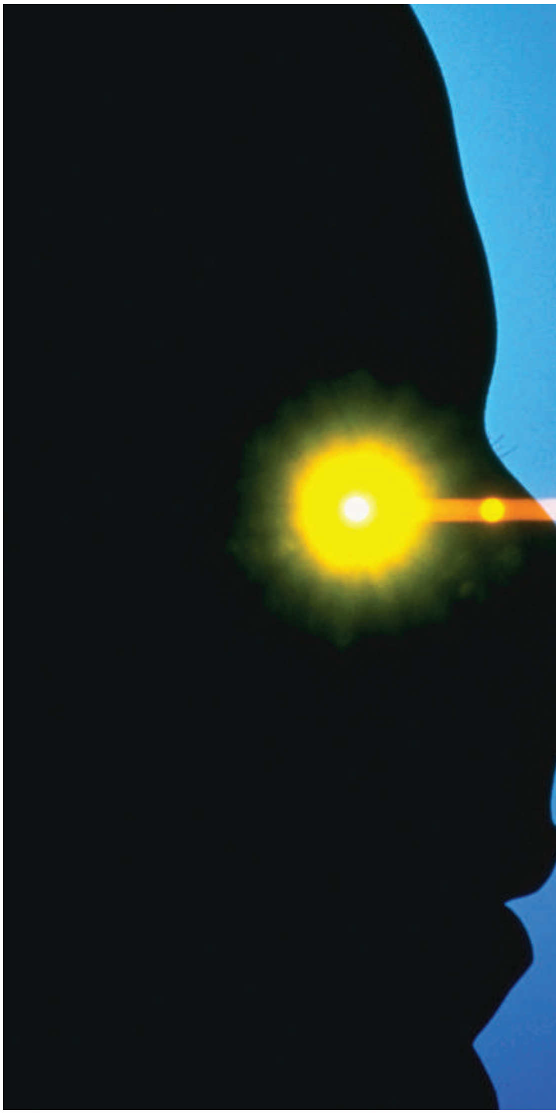
A recent experiment (M. Fine & D. Tchernov *Science* **315,** 1811; 2007) resolves this apparent paradox. The team grew scleractinian corals for a year in sea water with a lower-than-normal pH. They found that the corals reproduced and grew happily in this acidic environment — albeit without their hard skeletons. The corals adjusted their skeleton-forming physiology in response to the different growing conditions.

So corals seem to be quite adaptable. But I would like to know whether other calcifying organisms have such physiological versatility.

Moreover, we have to remember that although corals may survive in an ocean with a lower pH as sea-anemone-like organisms, they are currently major contributors to the intricate physical structure of coral reefs. What will be the future of these ecosystems if their calcium-carbonate scaffolding disappears? Will our grandchildren enjoy the spectacular beauty of these 'rainforests' of the ocean?

NEWS

Proteins make light work of nerve control



Light-activated proteins could provide a fresh perspective on how nerve cells work.

There were audible gasps and spontaneous applause at a neuroscience meeting in Salt Lake City, Utah, in February, when Ed Boyden described a protein that switches off nerve firing when activated by light. And when Karl Deisseroth told the fuller story of the protein, called NpHR and published in this week's *Nature*, at Cold Spring Harbor in New York late last month, there was talk of a revolution in neuroscience. It is perhaps no surprise that intellectual-property disputes are looming.

The revolution could consign electrodes — neuroscience's staple tools — to the trash after a century of faithful service. They would be replaced by genetically engineered proteins that allow investigators to stimulate or inhibit very precise groups of nerves at the flick of a light switch. No previous technology has come close to this level of control and precision.

"It is incredibly exciting — now we can really start to investigate how different neuronal cell types contribute to the neural circuits that mediate all sorts of behaviours," says Carl Petersen of the EPFL Brain Mind Institute in Lausanne, Switzerland. Petersen has already received the NpHR protein from Deisseroth's lab at Stanford University in California and is rushing to use it in his research on sensory perception. "It is the best thing that has happened in neuroscience in a good long time."

This feeling of urgency pervades the field. The technology is so powerful that leaps are predicted in many areas. With such prizes to be won, there is also a rush to publish. Boyden, a former postdoc of Deisseroth's who left Stanford shortly after the NpHR work began and is now at the Massachusetts Institute of Technology (MIT) in Cambridge, hurried through a report last month on the activity of NpHR in cultured brain cells (X. Han and E. S. Boyden *PLoS One* **2,** e299; 2007). Boyden says he thinks the idea belongs to him, but both MIT and Stanford are pursuing patents.

The *Nature* paper has resulted from a collaboration between researchers in Germany and at Stanford University. It extends the collaboration's 2005 work, conducted with Boyden, on a channel for positively charged ions (such as calcium) that is found in green algae and is activated by blue light. In that work, the researchers transplanted the channel, ChR2, into mammalian neurons. For the first time, it was possible to stimulate a nerve remotely at speeds close to normal neuronal transmission (E. S. Boyden *Nature Neurosci.* **8,** 1263–1268; 2005). Numerous research groups have already begun to use this 'on switch'.

The newly reported NpHR protein (see page 633) is exciting researchers even more. Identified in an archaeal species called *Natronomonas pharaonis*, it pumps chloride ions into cells, silencing physiological activity, when activated by yellow light. "The 'on switch' means we can replace the crude electrode, which stimulates all types of neurons in its vicinity," says Deisseroth. "But with the 'off switch' we can start to understand what is going on physiologically — or pathologically." By turning off sets of neurons in turn, researchers can investigate which ones are necessary, or sufficient, to elicit a particular behaviour or response.

Deisseroth and his colleagues have transferred the genes for NpHR and ChR2 into the nematode *Caenorhabditis elegans*, and can start the worms' swimming movements with flashes of blue light and stop them with yellow light. They also showed that functional proteins are produced when the genes are injected into the brains of young mice.

In addition to his research, Deisseroth holds a weekly psychiatric clinic, in which he assesses whether severely depressed patients are suitable for a treatment called deep-brain stimulation. In this procedure, electrodes are implanted deep in the brain to try to activate the neuronal circuits that lift mood. But the technique is crude and experimental, and Deisseroth says that the plight of his patients made him want to find something better.

"It is the best thing that has happened in neuroscience in a good long time."

The light-operated proteins might eventually replace electrodes in deep-brain stimulation, allowing physicians to hit just those neurons relevant to the disease being treated, although this would require a safe way to transfer the proteins into human brain cells. The technique could also have shorter-term clinical implications. For example, Gary Matthews of the State University of New York at Stony Brook hopes to use the switches to persuade retinal neurons, which don't respond directly to light, to mimic the responses elicited by rods and cones, to see whether this could help restore vision.

But the immediate use for the technology will be dissecting the role of different types of neurons in the circuits of both healthy and diseased brains. Deisseroth plans to use mice that express both proteins to identify targets relevant to depression, whereas Boyden plans work on mouse models of epilepsy, depression and Parkinson's disease.

Both researchers are distributing the NpHR protein to colleagues around the world, such as Sergey Kasparov at the University of Bristol, UK, who studies neurotransmitter release. When Kasparov heard about Deisseroth's work, he jettisoned a complicated plan to silence neurons that use noradrenaline as a transmitter: "The question we were posing is better answered by

RACE FOR A GREEN CAR
Multimillion-dollar X Prize set for automotive innovation.
www.nature.com/news

S. RAMELLA

ALAMY

the light-activated protein technology."

Another researcher keen to use the protein is David Kleinfeld of the University of California, San Diego, who is tracing the neuronal pathways that mediate touch sensations. "I moved very quickly to get a material-transfer agreement after we heard Deisseroth talk about the work," he says. "We are really psyched up about it."

But Petersen cautions that the intellectual-property issues surrounding such a significant technology "should be huge". So far, the parties involved are commenting little on the conflicting claims. Deisseroth points out that Boyden was supported by his Stanford lab when the work on NpHR began there. But both claims may have to fight their way round a 1991 patent awarded to Japanese scientists, which broadly covers light-activated channels. ■

Alison Abbott

Car emissions are EPA's problem

In a major victory for environmentalists, the US Supreme Court this week ruled that the federal government must regulate carbon dioxide emitted from cars and trucks.

In a strongly worded opinion issued on 2 April, the justices brushed aside insistence from the Environmental Protection Agency (EPA) that it did not have the authority under the Clean Air Act to regulate greenhouse-gas emissions from vehicles. They also disagreed with the agency's contention that even if it had the authority, it doesn't have to regulate if it doesn't want to.

"The use of the word 'judgment' is not a roving license to ignore the statutory text," wrote Justice John Stevens. "Under the clear terms of the Clean Air Act, EPA can avoid taking further action only if it determines that greenhouse gases do not contribute to climate change or if it provides some reasonable explanation as to why it cannot or will not exercise its discretion to determine whether they do."

Led by Massachusetts, the case was brought by several states, a handful of environmental groups and American Samoa. They argued that climate change was, among other things, causing the sea to rise up and swallow some states' territory and that the EPA was breaking the law by refusing to consider greenhouse gases as 'pollutants', which it has to regulate under the Clean Air Act.

The EPA denied this on a number of grounds, including appealing to the intentions of Congress when creating the act, and the desirability of using the promise of reducing emissions as a lever to force developing countries to do the same — something that wouldn't be possible if the EPA was already regulating them (see *Nature* 443, 486–487; 2006).

Lead author of the petition for Massachusetts, Lisa Heinzerling of Georgetown University Law Center in Washington DC, is naturally pleased with the outcome: "On every single issue we won, and I think we won big."

The ruling means that the EPA must now evaluate greenhouse gases and determine whether they are "air pollution which may reasonably be anticipated to endanger public health or welfare". If it finds that they are, it must regulate them. But there is no timeline for how quickly it must do this. The EPA under the Bush administration "will not move on it at all", says David Bookbinder, director of climate litigation at environmental group the Sierra Club, who was among the petitioners. "They have 18 months [left], they are a lame duck; they are not interested in climate-change regulation."

It is more likely that the next US president will begin to regulate greenhouse-gas emissions through the Clean Air Act, he says, if Congress doesn't get there first by passing climate-change laws. Certainly, many of the Democrats in Congress would love to regulate emissions in a similar manner to the European Union's Emission Trading Scheme.

Senator Barbara Boxer (Democrat, California), head of the Senate Committee on Environment and Public Works, plans to call EPA officials to Capitol Hill later this month to ask them how they will begin regulating climate change. In a statement, she said: "We now have a two-track process for addressing global warming — comprehensive legislation and administrative action."

The EPA is "reviewing the court's decision to determine the appropriate course of action", says spokeswoman Jennifer Wood, adding that "the Bush administration has an unparalleled financial, international and domestic commitment to reducing greenhouse-gas emissions" — a slightly bizarre contention, to say the least. ■

Emma Marris

The Environmental Protection Agency now faces the task of regulating greenhouse-gas emissions from vehicles.

M. BLAKE/REUTERS

China plans to modernize traditional medicine

BEIJING
China has announced an ambitious attempt to bring the ancient practice of traditional Chinese medicine (TCM) into line with modern standards. The government says it will expand basic and clinical research, and improve the testing and developing of TCM remedies for export. But critics question whether the research will meet the scientific standards necessary for international recognition.

Although China has invested in TCM before, the 15-year plan involves a change of direction and will receive significantly more money. The project is backed by 16 Chinese ministries, spearheaded by the Ministry of Science and Technology (MOST), the Ministry of Health and the State Administration of Traditional Chinese Medicine (SATCM).

Zou Jian-qiang, director of MOST's health and biotechnology division, says the government has earmarked 1 billion yuan (US$130 million) for TCM research and development, with the total budget over the next five years increasing to at least 5–6 times as much as the previous total of 740 million yuan.

In a country where most of the population is not covered by the public healthcare system and cannot afford to go to hospital, there is also a tenfold increase in money for the TCM-related part of the public healthcare system to 8.5 billion yuan. "Chinese medicine, which has served the Chinese people since antiquity, still has an important role in today's healthcare, especially in areas where people do not have access to, or could not afford, treatments based on Western medicine," says Yu Wen-ming, deputy director of SATCM.

The initiative comes after heated discussions on TCM throughout China. Last year, Zhang Gong-yao, a scholar at the Central South University in Changsha, Hunan, sparked a national debate when he published an article titled "Farewell to Traditional Chinese Medicine" in the Chinese journal *Medicine and Philosophy* (**27,** 14–17; 2006). Zhang argued that TCM is a pseudoscience and should not be part of public healthcare and research.

All sides of the argument have been keenly awaiting the government's stance, and it has come down firmly in favour of promoting TCM as a science. China is desperate to earn regulatory approval for TCM remedies abroad, and hopes to globalize TCM by 2020. The world market for Chinese herbal medicine has doubled over the past decade, with Europe and the United States being the biggest importers. But the patchy safety record of TCM from China has led to a steady decline in its exports, and it has lost market share to neighbouring nations such as Japan and South Korea.

So China's plan specifies strategies to boost standards, including conducting clinical research on the safety and efficacy of TCM remedies, encouraging international collaboration, improving manufacturing techniques and bringing the drug regulatory system into line with international guidelines.

More controversial is the government's shift in approach. Previous attempts to develop scientifically tested drugs from TCM have focused on isolating active ingredients from the remedies and screening them one at a time. This has led to a handful of new treatments, such as artemisinin for treating malaria and the decongestant ephedrine, but there has been no goldrush of approved drugs. The new plan aims to develop methodologies to test TCM's more traditional features and principles. The practice takes a holistic approach to disease treatment, so rather than using one drug to treat a particular disease, complex combinations of plant extracts are used. Each mixture is personalized to the symptoms and characteristics of the patient.

"The wind is now right for the development of traditional Chinese medicine."

TCM practitioners and researchers are still waiting to see whether the government will actually come up with the money, but they have generally welcomed the plan. "It's a significant step forward," says Liu Ping, vice-president of Shanghai University of Traditional Chinese Medicine. Some, however, are reluctant to jump on the bandwagon. A professor at the China Academy of Chinese Medical Sciences in Beijing, who asked not be named, reckons that after thousands of years of practice and development, TCM is already close to perfect and that modernization will simply distort its essence.

Shi-min Fang, a US-trained biochemist who now runs a website called 'New Threads' that fights pseudoscience and research misconduct in China, is also unimpressed by the plan, but for opposite reasons. He is in favour of scientific research into Chinese herbal remedies, but thinks the emphasis on testing the theories of TCM is misplaced. "The basic concepts of Chinese medicine, such as yin and yang, wu xing (the five elements) and the qi (meridian) theory, are inaccurate descriptions of the human body that verge on imaginative," he says. "The government has already spent a lot of money trying to prove their mechanistic basis, but this hasn't gone anywhere."

Wang Ming-wei, director of the National Centre for Drug Screening at the Shanghai Institute of Materia Medica, agrees: "To truly modernize Chinese medicine, we must first demystify these theories."

Some critics also worry that the plan doesn't set strict enough scientific standards. Although clinical research is listed as a priority, the plan doesn't specify whether there should be randomized, controlled trials in which neither practitioners nor patients know who is receiv-

YANG LIU/CORBIS

Traditional Chinese medicine uses mixtures of plant extracts in a holistic approach to disease.

ing active remedy and who is getting a placebo. And there is no requirement for TCM researchers to publish in internationally recognized journals. "Most research on TCM in the past is of poor quality, and is published only in Chinese medical journals without proper peer-review processes," remarks Wang. "Without a clear position from the government, it is unlikely that the situation will change."

Another concern is that the government does not specify how it will control the way in which research funding is allocated. Some critics feel that resources are currently being circulated only among certain favoured universities and institutes, and argue that reform to ensure that TCM grants are based on merit is necessary if any real progress is to be achieved. Otherwise, as Jia Wei, associate dean of the pharmacy school at Shanghai Jiaotong University, puts it, the initiative will be just "loud thunder, small raindrops".

Zou says that the government is aware of the controversy and problems surrounding TCM, and that the plan, on which it has consulted widely, is set to resolve these issues. Su Gangqiang, deputy director-general of SATCM's science and technology department, points out that the plan outlines overall principles and long-term directions for the development of TCM, rather than going into details, and says that further strategies will be published to specify exactly how the schemes will be carried out.

Whether the government will take critics' concerns into account while drafting these plans remains to be seen, but many are cautiously optimistic. "The wind is now right for the development of TCM," says Jia. "Let's hope this will not be a missed opportunity." ■

Jane Qiu

SIDELINES

SCORECARD

School projects
GlaxoSmithKline, which makes the blackcurrant soft drink Ribena, has been fined after a project by two New Zealand schoolgirls found that, contrary to the company's claims, the syrup for the cordial contained almost no vitamin C.

Russian satellites
Passengers on a Chilean flight to New Zealand were stunned when they were narrowly missed by flaming chunks of a plummeting Russian satellite, apparently removed from orbit at the wrong time.

OVERHYPED

Space marathons
It seems that you don't have to be in Boston to run the Boston Marathon these days — astronaut Sunita Williams is planning to run the race on a treadmill on the International Space Station. Presumably officials will discount the fact that she will already be travelling at more than five miles per second...

3 GOOD REASONS...

... to look at a webcam this week
The Internet's latest cult hit is www.cheddarvision.tv, which allows fans of cheese and/or mould to witness the maturation process live as it happens.

WEST COUNTRY FARMHOUSE CHEESEMAKERS

And if you don't do dairy, why not try some armchair astronomy at www.astro.caltech.edu/palomar/webcam.html.

Or be thankful you're warm and dry as you check the view from the Antarctic research vessel *RRS James Clark Ross* (www.antarctica.ac.uk/Living_and_Working/Transport/Ships/Webcam).

Sources: Reuters, thewest.com.au, PlanetArk, NASA

CAUDWELL XTREME EVEREST

EVEREST EXPEDITION
The team riding an exercise bike on Everest tells all on the *Nature* newsblog.
http://blogs.nature.com/news

Joan of Arc's relics exposed as forgery

PARIS
The relics of St Joan of Arc are not the remains of the fifteenth-century French heroine after all, according to European experts who have analysed the sacred scraps. Instead, they say the relics are a forgery, made from the remains of an Egyptian mummy.

Joan was burned at the stake in 1431 in Rouen, Normandy. The relics were discovered in 1867 in a jar in the attic of a Paris pharmacy, with the inscription "Remains found under the stake of Joan of Arc, virgin of Orleans". They were recognized by the Church, and are now housed in a museum in Chinon that belongs to the Archdiocese of Tours.

Philippe Charlier, a forensic scientist at Raymond Poincaré Hospital in Garches, near Paris, obtained permission to study the relics from the French church last year. He says he was "astonished" by the results. "I'd never have thought that it could be from a mummy."

Charlier and his colleagues didn't have much to work with: the relics comprise a charred-looking human rib, chunks of what seem to be carbonized wood, a 15-centimetre fragment of linen and a cat femur — consistent with the medieval practice of throwing black cats onto the pyre of supposed witches.

J. FOUCHET/SIPA/NEWSCOM

A vanilla smell of the alleged remains from Joan of Arc suggested natural decomposition, not burning.

Sniff tests

The researchers used a battery of techniques to investigate the remains, including mass, infrared and atomic-emission spectrometry, electron microscopy, pollen analysis and, unusually, the help of the leading 'noses' of the perfume industry: Sylvaine Delacourte from Guerlain, and Jean-Michel Duriez from Jean Patou.

Odour analysis is a new technique for palaeopathology, but Charlier says that he hit on the idea after being struck by the variety of odours of other historical corpses. Delacourte and Duriez sniffed the relics and nine other samples of bone and hair from Charlier's lab without being told what the samples were. They were also not allowed to confer. Both smelled hints of 'burnt plaster' and 'vanilla' in the samples from the relics.

"I'd never have thought that it could be from a mummy."

The plaster smell was consistent with the fact that Joan of Arc was burnt on a plaster stake, not a wooden one, to make the whole macabre spectacle last longer. But vanilla is inconsistent with cremation. "Vanillin is produced during decomposition of a body," says Charlier. "You would find it in a mummy, but not in someone who was burnt."

Other, more conventional, evidence pointing to a mummy origin quickly accumulated. Microscopic and chemical analysis of the black crust on the rib and on the cat femur showed that they were not in fact burnt, but were impregnated with a vegetal and mineral matrix, with no trace of muscle, skin, fat or hair. "I see burnt remains all the time in my job," says Charlier. "It was obviously not burnt tissue."

The black material was, however, consistent with an embalming mix of wood resins, bitumen and chemicals such as malachite. It was also consistent with gypsum, which gives the mix its plaster smell. The linen cloth had a coating characteristic of mummy wrappings. And large amounts of pine pollen were present. Pine trees did not grow in Normandy at the time that Joan of Arc was killed, but pine resin was used widely in Egypt during embalming.

Two other lines of evidence seem to clinch the mummy origin. Carbon-14 analysis dated the remains to between the third and sixth centuries BC. And the spectrometry profiles of the rib, femur and black chunks matched those from Egyptian mummies from the period, and not those of burnt bones.

Charlier points out that mummies were used in Europe during the Middle Ages in pharmaceutical remedies. The 1867 discovery date also fits the period when Joan of Arc, who had been forgotten for centuries, was rediscovered by historians and created as a national myth. Someone might have forged the relics at this time in an attempt to reinforce her importance.

"It is a fascinating project," says Anastasia Tsaliki, a palaeopathologist at the University of Durham, UK. Palaeopathology is a small but emerging field that attempts to use forensic science to inform history, traditionally a social science. "Philippe's work goes a step further by showing how forensic methods can be combined with tools used in archaeometry, archaeobotany and osteology," says Tsaliki.

Fire-proof organs

Part of the legend of Joan of Arc springs from the observation, documented in historical records, that some of her organs resisted the fire. Hundreds of pages of surviving manuscripts describe in vivid detail how she was burnt three times over to try to ensure that nothing but ash remained, and so prevent her remains being worshipped. The observation of remaining organs was interpreted as a miracle.

But science has another explanation. "In fact, it is very difficult to totally cremate a body; organs such as the heart and intestines, which have a high water content, are very resistant to fire," says Charlier. "We see it all the time in forensics."

Debunking the relics of Joan of Arc will be less controversial than doing the same for the Shroud of Turin, but is still likely to generate large public interest, especially in France. The Church is ready to accept the results, according to Charlier. ■

Declan Butler

Smithsonian looks beyond ousted boss

By any standards, it was a controversial reign. It started with an investigation into the purchase of feathers from endangered birds; not what you might expect from a man with ultimate responsibility for one of the world's most prestigious natural history museums. It ended last week amid media claims of lavish expense accounts, including thousands of dollars for the upkeep of a private swimming pool and the cleaning of a chandelier. It is unlikely that many tears were shed at the Smithsonian Institution on 26 March, when officials announced that Larry Small had resigned.

The departure of Small, who has led the 19 museums and 9 research centres that make up the institution since 2000, closes one of the more turbulent periods in the organization's 160-year history. Yet researchers at the institution say that the problems they face run deeper than those allegedly caused by Small's management approach (see 'A lavish seven years').

The Smithsonian's core research budgets have been falling for 15 years.

L. SNIDER/PHOTO IMAGES/CORBIS

For around 15 years, the Smithsonian's core research budgets have eroded, forcing administrators to rely increasingly on outside funds. And although donors can often be persuaded to stump up for glamorous projects such as new exhibitions, finding funds for research costs is much more challenging. As a result, although Small carried out much-needed repairs and launched new museums, he did not stop the slide in funding for research and training. Of the nearly US$1 billion he raised in private money, almost all went to infrastructure projects such as new galleries, rather than to research.

"This issue won't go away because Larry has gone," warns Stephen Murray, deputy director for science at the Harvard-Smithsonian Center for Astrophysics in Cambridge, Massachusetts.

Smithsonian officials say that they don't have figures for how much money from the institution's overall budget goes towards science, but research leaders think that the attrition in funding has hit many areas. Researchers from a variety of specialities told *Nature* that as many as a third of full-time positions have been lost from their departments over the past 15 or so years. Where centres have prospered, it has generally been because they tapped into funding streams that became national priorities, such as money for research on climate change and its effects, or could raise money from philanthropic sources.

Although researchers think that a mix of funding streams is a good thing, they don't want this strategy to replace core federal funding. For example, the Smithsonian's centres hold records on forest growth and bird migration and behaviour that span 25 years. Such long-term data — one of the institution's great strengths — are hard to maintain with agency grants, which generally run for only three to five years.

"The value of long-term monitoring depends on its continuity," says Ira Rubinoff, acting under-secretary for science at the institution. Rubinoff, has secured a private donation to maintain the forest work, but only for five more years. Like other researchers, he says that he now needs to devote more time to fundraising.

With this history in mind, Smithsonian scientists say that beefing up core funding should be a priority for the next director. That feeling is shared by the new acting head, Cristián Samper, director of the National Museum of Natural History in Washington DC, where core funding has decreased by around 3% a year over the past eight years. Samper hopes that the trend can be reversed by improving the advertising of Smithsonian research to Congress and the public, in part by getting more of that science into the exhibitions that the Smithsonian runs.

Samper is one of many names being suggested in discussions over Small's successor, but he comments only that he would be happy to remain as head permanently, or to go back to running the national museum. The search is being run by members of the museum's Board of Regents and is expected to take about a year.

Jim Giles

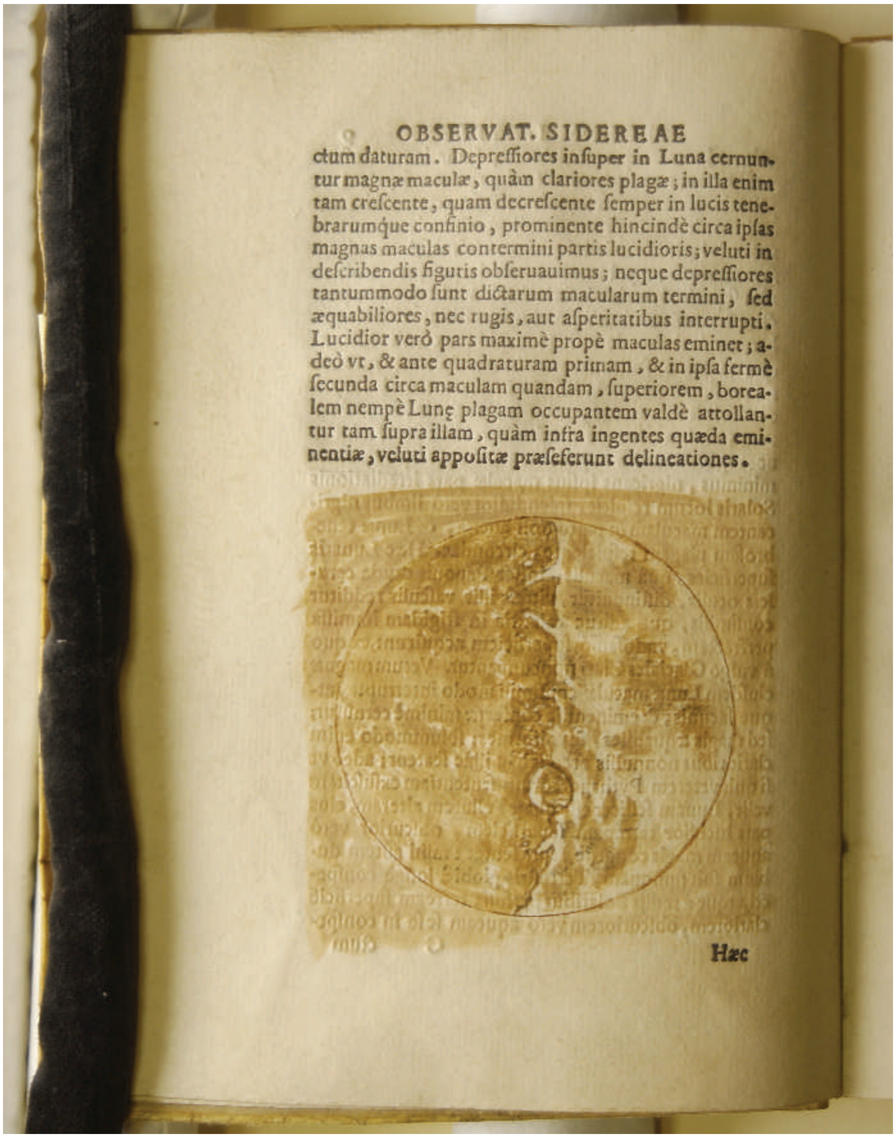
A lavish seven years

Larry Small was always likely to make waves at the Smithsonian Institution. As a successful business leader, he brought with him a corporate, goal-orientated attitude that is at odds with the academic freedom that scientists seek.

That approach brought benefits: an ability to raise funds that led to the creation of new museums devoted to American Indians and aerospace, for example. But Small also angered scientists by proposing in 2001 to close two research centres without first running detailed consultations. Opposition from staff and Congress forced him to back-track.

Earlier that year, Small had been investigated for his collection of Latin American masks, head-dresses and costumes. He was eventually prosecuted in 2004 for purchasing artefacts that contained feathers from endangered birds and served 100 hours' community service.

But it was Small's expense account that caused most anger in Congress and eventually forced him out. He used his Washington home for Smithsonian functions and allegedly received more than $1 million in housing allowances during his seven years in charge, according to *The Washington Post*. The newspaper claims that invoices included $273,000 in housekeeping services, $2,535 to clean a chandelier and $12,000 for the upkeep of his swimming pool. The expenses were permitted under Smithsonian policy, although an internal institution report, issued in January, had noted that some of Small's expenses "might be considered lavish or extravagant". When he quit last week, Small was earning more than $915,000. **J.G.**



OBSERVAT. SIDEREAE

ctum daturam. Depreſſiores inſuper in Luna cernuntur magnæ maculæ, quàm clariores plagæ; in illa enim tam creſcente, quam decreſcente ſemper in lucis tenebrarumque confinio, prominente hincindè circa ipſas magnas maculas contermini partis lucidioris; veluti in deſcribendis figuris obſeruauimus; neque depreſſiores tantummodo ſunt dictarum macularum termini, ſed æquabiliores, nec rugis, aut aſperitatibus interrupti. Lucidior verò pars maximè propè maculas eminet; adeò vt, & ante quadraturam primam, & in ipſa fermè ſecunda circa maculam quandam, ſuperiorem, borealem nempè Lunę plagam occupantem valdè attollantur tam ſupra illam, quàm infra ingentes quæda eminentiæ, veluti appoſitæ præſeferunt delineationes.

Hæc

B. HERRENKIND

SNAPSHOT
Over the Moon

Galileo's education included artistic training but his soul was that of a modern scientist: he snapped up new technologies, and was paranoid about being scooped.

When he realized, while pointing his prototype telescope at the Paduan sky, that the Moon was not the smooth, pristine sphere that religious leaders would have us believe, he rushed to publication. His artistic training, reflecting a new awareness of perspective, helped him to understand that the dark spots were shadows thrown by an uneven surface of mountains and craters, and to render this beautifully.

His last observation in his book *Sidereus Nuncius* (*The Starry Messenger*) was on 2 March 1610 — one risky day after the Holy Inquisition had approved the text for publication — and 550 copies of the book were ready for distribution to his peers on 12 March. Johannes Kepler received his copy in Prague on 8 April.

Only 500 of those copies included the engravings made from Galileo's famous sketches of the Moon. Now, previously unseen paintings have surfaced in one of the remaining copies, recently acquired in South America by an antiquarian — no doubt a happy man. Last week, art and science historians formally authenticated the five yellowy paintings as being from the hand of Galileo.

Alison Abbott

Magnet fails initial test at Large Hadron Collider

Construction of the Large Hadron Collider (LHC) hit a snag on 27 March when one of the device's magnets underwent a 'serious failure' during testing.

Staff at the particle collider, which is based at CERN, near Geneva, Switzerland, were running checks on the superconducting magnets that will focus its proton beam. By applying a force on the magnets, they simulated the loss of superconductivity in the device, which would cause a rapid rise in temperature and put a huge strain on the cooling circuits used to keep the magnets close to absolute zero. But one of the structures holding a magnet broke when the pressure reached 20 atmospheres.

Officials at Fermilab, the US particle-physics lab that designed and built the magnets, say that the structure had not been built to specifications required by LHC scientists. A Fermilab team is working on a solution and says it hopes that the LHC launch in summer 2008 will not be delayed.

Indonesia lifts blockade on access to bird flu samples

Indonesia last week agreed to resume sending samples of the H5N1 avian flu virus to the World Health Organization (WHO). Such samples are needed for the agency to monitor how the virus is evolving and so to aid production of an effective vaccine.

Indonesia had stopped sending samples in January, protesting that its compliance should in return guarantee it access to affordable vaccines in the event of a pandemic. Instead, the country had struck a commercial deal with the US firm Baxter Healthcare in which Indonesia would get help in building vaccine plants in return for providing virus samples.

Under the latest agreement, which was reached at a meeting in Jakarta, the WHO will get samples and will be allowed to share them with researchers. However, it

I. FEDRIANSYAH/AP

Indonesia hopes to secure vaccines against bird flu.

will only be able to share the samples with commercial firms if the companies have already negotiated access with Indonesia.

More people have died of H5N1 in Indonesia than anywhere else in the world.

Iranians are dropped from American Chemical Society

Thirty-six Iranian researchers have had their membership of the American Chemical Society (ACS) cancelled because the society believes it has broken federal law by dealing with them.

The United States has imposed a total embargo on Iran, although professional societies still tend to take Iranian members because the measure does not to extend to scholarly materials. But ACS assistant general counsel David Smorodin reviewed the embargo rules recently and decided that selling its journals at a discounted rate — a benefit the ACS offers to all members — would be seen as a commercial transaction that went beyond this exception. The Iranian members were then told that their memberships had been cancelled.

The ACS says it plans to apply to the Department of Commerce for a licence that exempts it from the embargo and that it hopes to be able to reinstate the Iranian researchers.

Japan's next supercomputer finds a home in Kobe

Japan is trying to regain its title as possessor of the world's fastest supercomputer, and last week it revealed that the machine will be housed in Kobe, which is aiming to become a major medical hub.

Last year, engineers began drawing up specifications for a computer that can operate at a speed of 10 petaflops, or 10 quadrillion calculations per second. That will make it more than 250 times faster than Japan's Earth Simulator, once the world's fastest supercomputer but currently ranked 14th. The main application for the new Japanese machine will be nanotechnology and life-science research, such as protein analysis, in addition to geoscience and astrophysics.

The ¥115-billion (US$1-billion) national project is expected to be completed by 2012.

PS: when not playing games please do some science

Video-gamers using Sony's Internet-enabled PlayStation 3 (PS3) console have been given the option to contribute to science.

More than 50,000 people have signed up to take part in a distributed-computing project called Folding@home, which uses the

K. KASAHARA/AP

Game on: PlayStation owners can now help science.

spare processing power of home computers to compute the way in which different proteins fold, in a bid to tackle diseases such as Alzheimer's. Project scientists at Stanford University partnered with Sony, whose PS3 has computing power equivalent to ten average desktop computers. Other distributed-computing projects may also soon be available on the PS3.

Currently, the program must run in the foreground. Sony is considering making it a background application, but it will not be able to work during game play, which uses too much processing power. So sadly no one will be able to claim they are contributing to science by playing *Grand Theft Auto*.

Russian academy resists state move to gain control

Members of the Russian Academy of Sciences have voted almost unanimously against a government-proposed charter that would transfer control of the academy to the state.

The 1,200-member academy, which oversees a budget of US$1.2 billion, some 400 research institutes and about 200,000 researchers and staff, has enjoyed relative autonomy for nearly 300 years. But last year, steps towards imposing government control began when the parliament passed a law stipulating that the academy's president must be approved by the government.

The new charter, prepared by the Ministry of Education and Science, calls for the creation of a nine-member supervisory board — consisting mainly of government representatives — to oversee the academy's work, budget and property. To be valid, the charter agreed by the academy would have to be ratified by the government.

Clarification

The News story "Congress requests bubble-fusion reports" (*Nature* **446,** 480; 2007) incorrectly implied that highly publicized hearings on allegations against the biologist Thereza Imanishi-Kari were held by the Investigations and Oversight Subcommittee of the House Committee on Science and Technology. The hearings were actually held by the House Committee on Energy and Commerce's Subcommittee on Oversight and Investigations.

BUSINESS

China's deadly drug problem

The head of China's drug-safety agency is under investigation for alleged corruption. **David Cyranoski** looks at how the inquiry might affect the country's fast-growing pharmaceutical industry.

A corruption scandal at China's top drug-regulatory agency has led to the arrest of its former head and an investigation of unprecedented scale into drugs approved during his tenure. Some are calling the crackdown a step towards a modernized drug industry that encourages innovation. But others wonder whether anything will really change. "The government is cleaning up, but many hope it will be changed once and for all," says Wang Ming Wei, director of the National Center for Drug Screening in Shanghai. "You can't just do cosmetic surgery."

Faulty drugs: the antibiotic clindamycin has caused multiple deaths in China.

EYEPRESS/NEWSCOM

The scandal centres around Zheng Xiaoyu, who was the head of the State Food and Drug Administration (SFDA) from 1998 until June 2005. Last summer, concerns mounted about the agency after poor-quality drugs resulted in some highly publicized deaths. More than ten people died across the country after being injected with the antibiotic clindamycin, which had not been sterilized properly. And 11 people in southern China died from kidney failure after receiving the drug armillarisni A, which had been manufactured with a toxic substance.

A high-level probe of the SFDA culminated in Zheng being detained in December 2006. In late January, the Chinese premier Wen Jiabao ordered an investigation into reports that Zheng had accepted bribes, and Zheng was subsequently thrown out of the Communist Party. Two of Zheng's former secretaries have also been detained. Hao Heping, who headed the department that bestowed licences to manufacture medical equipment, was sentenced to 15 years in prison for accepting money to smooth drug approvals. Cao Wenzhuang, former head of the drug-registry division, has also been detained.

> "Many hope the drug industry will be changed once and for all. You can't just do cosmetic surgery."
> — Wang Ming Wei

Zheng's wife and son were also arrested, according to Chinese news reports. A senior executive of a foreign pharmaceutical company who did not want to be named says that they ran consulting firms whose work would help ensure a drug's approval.

China's US$13-billion pharmaceutical industry has struggled to overcome deeply rooted problems. Many drug companies in China make money not through innovation but by making different formulations of products. Statistics provided by Wang show that about half of the 'new' drugs are varieties of previously approved products. Yet large numbers of applications flood the SFDA; in 2004, for instance, China fielded some 17,000 applications for new drug approvals, compared with 167 for the US Food and Drug Administration.

At first, Zheng seemed like a reformer who was trying to modernize the country's regulatory oversight. A veteran of the drug industry, he led a merger of the state food administration and the state drug administration — which he had headed — into the SFDA in 1998. He introduced a Chinese version of Good Manufacturing Practice guidelines, which ensure that production is up to standard and that a mechanism is in place to deal with drug problems. All Chinese drug companies had to comply with the requirement by July 2004; more than a quarter closed their doors. Zheng also established new requirements that traditional Chinese medicines had to fall in line with national, rather than previously established regional, standards.

Under the new pressures, observers say, drug companies became desperate to get their products approved by SFDA officials, who were able to use their positions to push drugs through the system. "There was too much power invested in a few individuals," says Wang. Government officials seem to agree. In February, Wu Yi, one of China's vice-premiers, reportedly told members of the SFDA: "We urgently need to reform the system to remove loopholes in food and drug administration and set up an effective mechanism for clean governance." She suggested several measures that could be implemented, such as requiring that the acceptance, appraisal and approval of new drug applications are each handled by different people.

Stifling standards

But as the SFDA tries to raise standards for its drug-approval processes, some are worried that smaller companies will not be able to survive. According to one Shanghai-based researcher involved in drug development, some clinical trials now require thousands of patients rather than the hundreds needed previously. The source blames "arbitrarily high" standards that came as a reaction from "demoralized" SFDA staff. "Because of the overreaction, people without deep pockets will suffer," he says.

EYEPRESS/NEWSCOM

David Liu, vice-president of Beijing Pharmaceutical, a wholesaler for drugs and medical devices, agrees that the system is geared towards taking smaller manufacturers out of the market. But it is a good thing, he says, because most of those companies do not innovate. "It is not like the United States, where you have small companies with an entrepreneurial spirit. In China, a small company means quick in, quick out," he says.

Multinational pharmaceutical companies that work in China could benefit. "A cleaned-up process will be good for foreign companies," says Liu. "They usually pay more attention to ethics in the first place." Several major drug companies, including Roche of Switzerland, AstraZeneca of the United Kingdom, and Eli Lilly of the United States, recently opened research and development facilities in China. And Novartis of Switzerland announced last November that it would invest US$100 million in a 400-researcher facility in Shanghai.

Private preferences

Senior executives from several multinational drug companies told *Nature* privately that they would be happy to see reforms in China's drug-regulatory system. "The cleaner, the better for us," said one. But when asked whether they had been encouraged to pay bribes or 'consulting fees' when applying to register drugs, or whether the SFDA scandal might affect their willingness to invest in the country, most shied away from official answers.

Within China, the fallout continues. The bribery scandal has pushed the agency to review the applications for some 170,000 drugs, focusing on those approved between 1999 and 2002. Sources in the pharmaceutical industry there say that SFDA officials and pharmaceutical companies are being given a grace period during which they can come clean and face a lesser penalty. The agency is also stepping up its investigation of companies and pharmacies thought to be involved in the scandal; 160 of these lost their licence to make or sell drugs in 2006.

There remains some scepticism as to whether the reforms will have a lasting effect. The inability for people — including most of the sources for this article — to voice their concerns suggests that the government might not have the input it needs. Some have even suggested that the reforms have been a political move and that things might return to their past state once officials under Zheng are removed. "You need another body to watch the SFDA," says Liu. Still, he, like most, is upbeat about the government's desire to reform. "The government knows it has a problem. Now it is going to fix it." ■

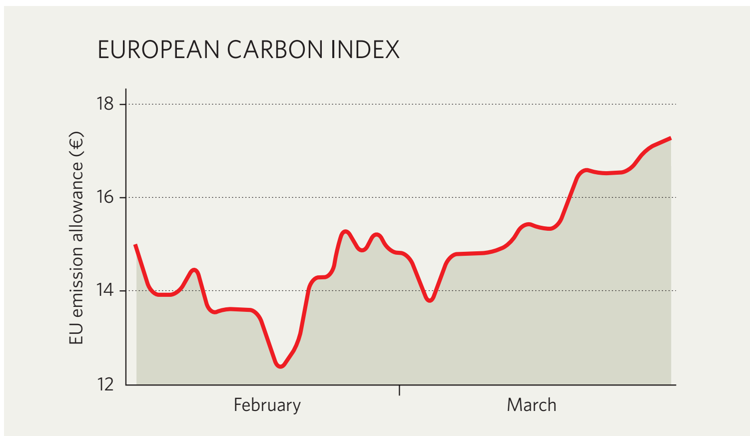
IN BRIEF

TROUBLED REPORTS **Indian scientist Raghunath Mashelkar, former head of the Council of Scientific and Industrial Research, has resigned as the chairman of a government panel on patent laws after charges that some parts of a recent report had been plagiarized. Non-governmental organizations, local pharmaceutical companies and some scientists had alleged that the panel allowed companies to monopolize drugs by letting them patent minor modifications to existing drugs. Critics revealed that key recommendations in the report had been lifted verbatim from an earlier study sponsored by US and European drug firms. Mashelkar had offered to correct the "inadvertent error" and resubmit the report, but critics want the entire report to be scrapped.**

TESTING TIMES **Two researchers who left the US government in the wake of Congressional hearings in 2002 on their work with private companies have founded a new firm to specialize in personalized medicine. Lance Liotta and Emanuel Petricoin say that their new company, Theranostics Health of Rockville, Maryland, will sell protein profiling tests that predict how patients' cancers will respond to treatments. The pair had previously developed a protein profiling test to detect ovarian cancer, but the test's accuracy was questioned by other scientists (see *Nature* 429, 496; 2004).**

NOT SO CLEAN **Torcetrapib, a drug abruptly dropped from clinical trials last December over safety concerns, might not be able to unclog arteries after all, according to studies reported last week at the American College of Cardiology meeting in New Orleans. Some of the studies have also been published in the *New England Journal of Medicine*. Pfizer had hoped that its drug would raise high-density lipoprotein, or 'good' cholesterol, enough to clear out arteries. The new findings suggest that good cholesterol does rise significantly, but has little effect on cleaning out arteries, leaving experts to wonder whether similar drugs under development will work. A much larger study whose expected release at the meeting was delayed (see *Nature* 446, 363; 2007) may answer that question.**

MARKET WATCH



SOURCE: EEX

The price of carbon allowances for phase two of the European Union emissions-trading system reached a three-month high last week, after the European Commission decided to cut Poland and the Czech Republic's plans to allocate allowances for carbon-dioxide emissions for the 2008–2012 trading period.

The commission will now permit Poland to hand out allowances of 208.5 million tonnes per year to its installations, which is 27% less than it asked for. The approved annual allocation to the Czech Republic is 86.8 million tonnes — 15% less than requested. France's national allocation plan — 132.8 million tonnes — was approved without cuts.

Power plants and other large installations can buy emission allowances on the market if they want to exceed their CO_2 caps. An allowance to emit one extra tonne of CO_2 during 2007 costs little more than €1 (US$ 1.33) because most installations own more allowances than they will need. But given the reduced caps for the second trading period, carbon futures for 2008 are now being traded at more than €17 — up from a low of €12 in late February.

"As from next year, there will likely be shortages," says Stefan Kleeberg, who watches carbon markets for 3C Climate Change Consulting near Frankfurt, Germany. "As a result, phase-two allowances could sell for €20–50 in the next months." ■

Quirin Schiermeier

A long time ago, in a galaxy not so far away

Vast stellar nurseries, clouds that dwarf the Solar System and lurking swarms of black holes. **Jeff Kanipe** probes the unfolding mysteries at the heart of the Milky Way.

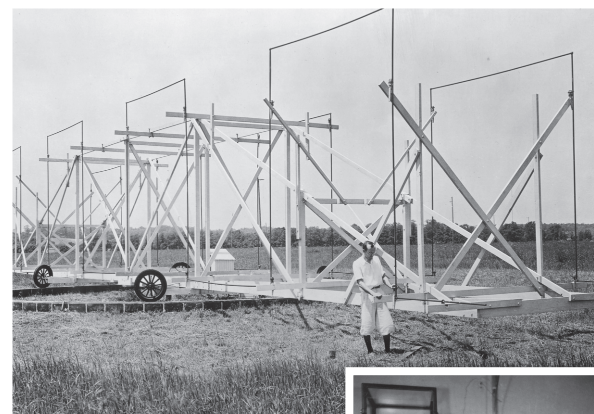
Before it was seen, it was heard. In the early 1930s, a Bell Labs engineer named Karl Jansky was given the job of sorting out where the static interference in radio transmissions came from. With an ungainly but ingenious steerable antenna he tracked a number of sources. Most were thunderstorms, but one wasn't. As Jansky tracked it across the sky from day to day he realized that it was far beyond Earth's atmosphere, and indeed beyond the Solar System — an abiding hiss from somewhere in the constellation of Sagittarius[1].

As constellations go, Sagittarius is modest both in size and in brightness. What sets it apart, on dark, moonless nights, are its background contrasts: brilliant, billowy clouds of stars that are punctuated by dusky rifts and voids. No other place in the sky looks this compelling.

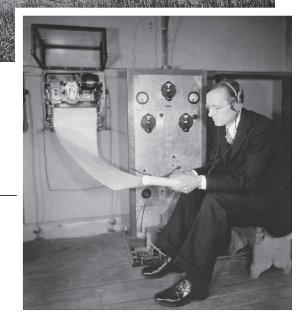
By the time of Jansky's discovery, the behaviour of other objects in the sky had already provided good evidence that something special lay within those beguiling clouds. In 1918, a study of star clusters by Harlow Shapley, an astronomer at the Mount Wilson observatory above Los Angeles, showed that 'open' star clusters were spread throughout the plain of the Milky Way, whereas globular clusters were concentrated in the direction of Sagittarius[2] — some above the Milky Way, and some below it, drawn by some unseen immensity. The globular clusters were like moths batting about a lamp hidden in Sagittarius's dense folds of dust.

Hidden behind the clouds

Jansky's observations provided the first hint of what lay behind those shrouds, but it took decades for further details to become clear. It was not until 1968 that the radio source at the centre of the Galaxy, now called Sgr A* (or 'Sagittarius A-star'), was detected in the infrared, showing that it was 1,000 times brighter than the radio emission had led astronomers to suspect[3]. At shorter infrared wavelengths — which like their longer brethren pass through dust much more easily than visible light — it was even brighter. By this time, astronomers gazing elsewhere in the sky had discovered quasars, bodies so bright and yet so small that it seemed possible they were powered by vast black holes sucking up dust and gas at an incredible rate. In 1969, the British astronomer Donald Lynden-Bell suggested that our Galaxy



LUCENT TECHNOLOGIES' BELL LABS



BETTMANN/CORBIS

Karl Jansky (right) used a rotating antenna to detect radio waves in the Milky Way.

and its neighbours could all have 'dead quasars' at their hearts[4]. By this line of reasoning, Sgr A* had to be a vast black hole.

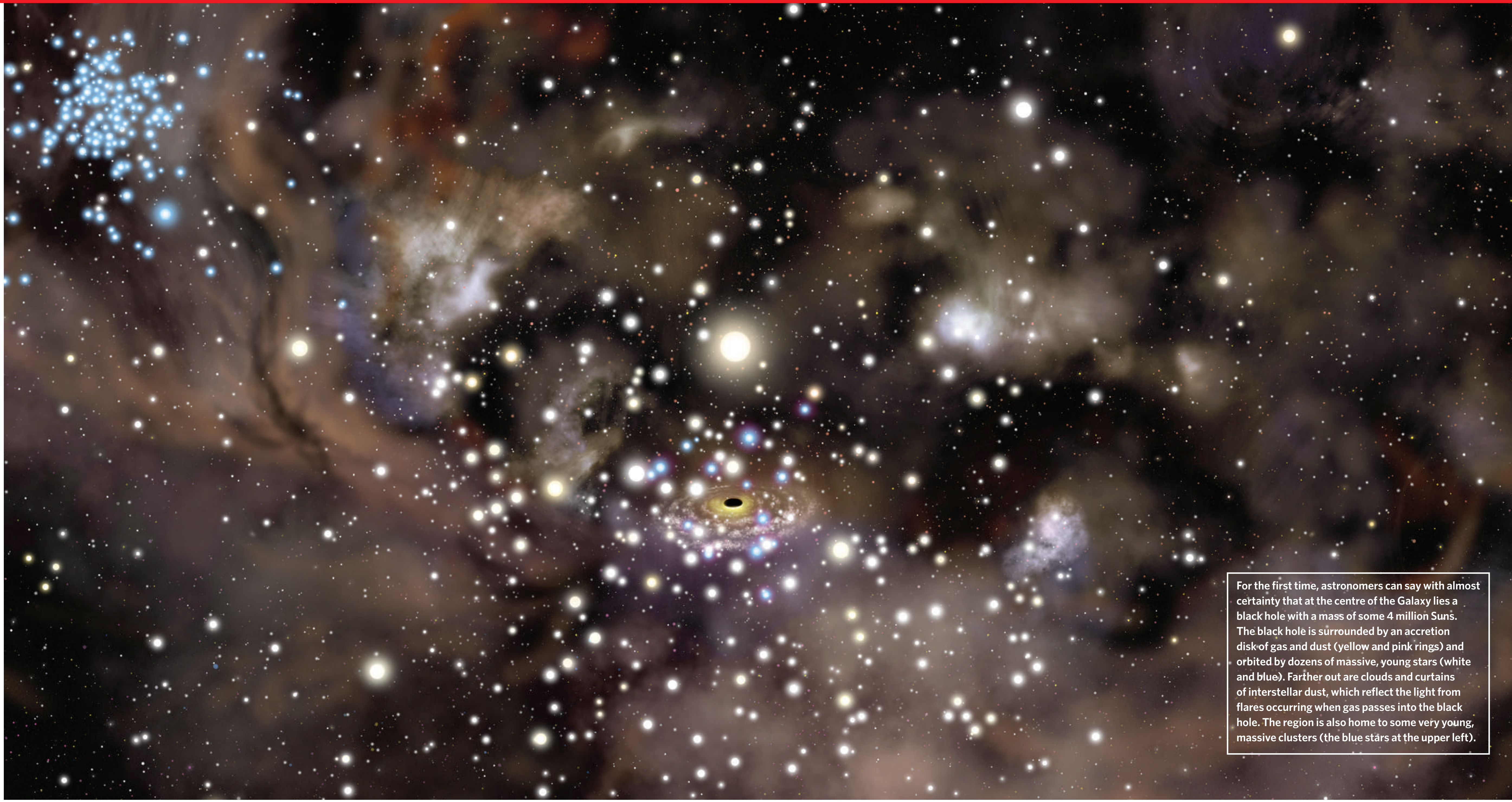
Even with a compelling theory, seeing the details was hard. Telescopes using visible light — even the mighty Hubble — could not see through the clouds of dust. But in the past ten years, powerful radio arrays, new infrared and X-ray telescopes, detectors in orbit and adaptive-optics systems on Earth have revealed strange new structures in and around the Galaxy's central engine: magnetic arcs and filaments, giant clumps of massive stars and whorls of gas. Analysis of the motions and masses of the stars within the central two light years of Sgr A* have shrunk the known heart of our Galaxy down to a region of space no larger than Earth's distance from the Sun, and probably much smaller, containing the mass of four million Suns. For all this insight, the heart of the Galaxy still has mysteries for astronomers to explore — from echoes of outbursts to the silhouette of the event horizon itself.

One problem with observing the very centre of the Galaxy, though, is that it simply isn't very bright — a firefly to the searchlight of a full-blown quasar. The obvious explanation for its lack of luminosity, says Andrea Ghez, principal investigator of the Galactic Center Group at the University of California, Los Angeles, is that even though Sgr A* lies at the heart of a galaxy of hundreds of billions of stars, it may be a bit isolated. The radiation from black holes comes not from the holes themselves, but from matter falling onto the accretion disks that swirl around them. There may just not be much matter around to fall onto the accretion disk at the centre of the Galaxy. Or the disk may be generating a wind of radiation strong enough to stop any more gas and dust flowing into it.

Unexpected echo

Nevertheless, the flow sometimes increases and the central engine heats up. For a couple of years in the 1950s, for example, Sgr A* looked perhaps a hundred thousand times brighter in the X-ray spectrum than it is today, probably because it swallowed a planet's worth of gas in a gulp. Unfortunately, humankind had no X-ray telescopes in the 1950s — the devices only work in space — which might seem a serious impediment to learning from the event. But it is not, it turns out, an insurmountable one.

At the January 2007 meeting of the American Astronomical Society in Seattle, Washington, Michael Muno of the California Institute of Technology (Caltech) in Pasadena announced that his team had managed to see part of the 1950s outburst reflected off clouds on the far side of the Galactic Centre[5]. X-rays that had started off heading away from Earth had bounced back to us, and because the clouds were a few tens of light years away from the centre, the reflected X-rays took half a century longer to get to Earth than did those that had taken the direct route. Half a century



ARTWORK BY LYNETTE COOK

For the first time, astronomers can say with almost certainty that at the centre of the Galaxy lies a black hole with a mass of some 4 million Suns. The black hole is surrounded by an accretion disk of gas and dust (yellow and pink rings) and orbited by dozens of massive, young stars (white and blue). Farther out are clouds and curtains of interstellar dust, which reflect the light from flares occurring when gas passes into the black hole. The region is also home to some very young, massive clusters (the blue stars at the upper left).

isn't much in terms of a journey across the 26,000 light years that separate Earth from the Galactic Centre, but it's enough to make the difference between astronomers stuck underneath Earth's X-ray-absorbing atmosphere and astronomers who, like Muno's team, can use NASA's Chandra X-ray telescope to monitor the centre of the Galaxy.

"This is the first X-ray echo that we have seen propagating through space after an event that we had not originally seen," says Muno. The observations allowed his team to say that the burst must have been 1,000 times brighter and 1,000 times longer than the contemporary ones seen with the Chandra telescope or the Japanese Advanced Satellite for Cosmology and Astrophysics. The intermittency of such events could imply that the disk of material swirling about the black hole is both meagre and unstable, only occasionally dropping a gobbet of matter into the black hole's maw.

But if the black hole's neighbourhood is by and large empty, how can we account for the family of bright young stars that swarms about it? This apparent paradox received prominent attention at the Galactic Center Workshop held in Bad Honnef, Germany, in 2006. The black hole's inactivity suggests that the central few light years doesn't contain enough raw material to make stars. And the enormous gravitational tidal forces around the black hole would seem to prohibit stars from forming even if the material were there: it's hard for a cloud of gas to contract into a star under its own gravity when something that weighs as much as four million stars is sitting next door.

Nevertheless, says Ghez, at the Galaxy's core is a swarm of about 40 massive young stars; they are called 'S stars' because they belong to the Sgr A* cluster. One, called S0-2, has a mass that is some 15 times that of the Sun and orbits Sgr A* with a period of just over 15 years. At its closest, it comes within 17 light hours of the supermassive black hole[6] — as close as the edge of our Solar System is to Earth. After 12 years of monitoring the motions of these stars using the infrared capabilities and adaptive optics of the W. M. Keck Observatory on Mauna Kea, Hawaii, Ghez's Galactic Center Group, has almost seen them make complete circuits of the centre and return to where they started: "We should see S0-2 close [its orbit] in 2010," she says.

The orbits of the central stars of the Galaxy can be used to further refine the mass of the central black hole and to constrain the distribution of mass in the neighbourhood. And their motions might also reveal something about how they got there in the first place.

There are two explanations for the stars' presence. One theory is that the stars formed more or less where they are today, near the black hole. In principle, this could have occurred if the density of the gases in the centre of the Galaxy was much higher in the past. Higher density would allow clumps in the clouds to collapse to form stars, even in the presence of a strong gravitational field.

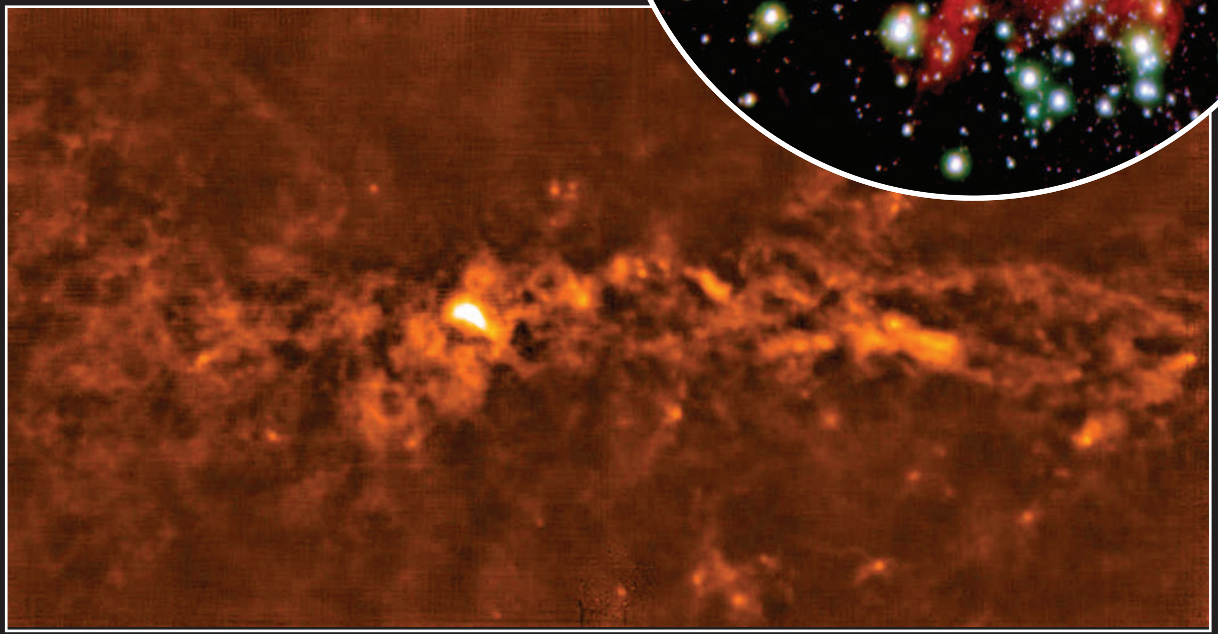
The alternative explanation is that the stars formed outside the adverse conditions of the central region and migrated there later on as a single massive cluster. However, for this to work the core of the original cluster would have needed a mass that was ten million times greater than that of the Sun, packed into a volume of no greater than 3 light years, which is more compact than any cluster known. At the moment, most astronomers seem to favour the first scenario.

But although young stars may not be migrating into the central zone, very old ones probably are. Theorists at the Galactic Center Workshop described recent simulations that bolster a striking prediction first made by Mark Morris of the University of California, Los Angeles, in 1993. Morris postulated that the inner three light years of the Galaxy's centre might contain as many as 20,000 star-sized black holes. These are the remnants of previous generations of bright young stars, which have sunk slowly in towards the central and much larger black hole over

J. CHUMACK/SPL

ESO

J. BALLY & THE BOLOCAM/CSO GALACTIC PLANE SURVEY TEAM



At the centre of the Milky Way (above) is a supermassive black hole called Sgr A* (right), which is surrounded by both hot (blue) and cold (red) stars. The stars form in 'nurseries' (below) as bright cores embedded in giant molecular clouds.

billions of years. The presence of a close-knit cluster of dead stars is supported by Chandra's discovery of four bright but variable X-ray sources — within 3 light years of Sgr A* (ref. 7). The sources' variability is a characteristic of systems in which matter from a normal star is sucked onto a black hole or an ultradense neutron star. Four fairly easily discerned X-ray sources of this type in such a confined region, say astronomers, provide strong circumstantial evidence that tens of thousands of black holes and neutron stars have settled in and around Sgr A*.

Smouldering stars

To see some real clusters, we have to step back a little. A hundred light years from Sgr A* lie the Arches and Quintuplet clusters — two of the most massive young clusters and unlike either the open or globular clusters seen elsewhere. The stars in the Arches cluster are 50 times closer together than are those in the neighbourhood of our Solar System. At the density seen in the Arches cluster, the space between the Sun and its nearest neighbour would contain 100,000 stars. The other cluster, the Quintuplet, is a bit older and more dispersed, but it has one of the biggest and potentially most volatile stars known, called the pistol star (so named for the pistol-shaped nebula in which it lies).

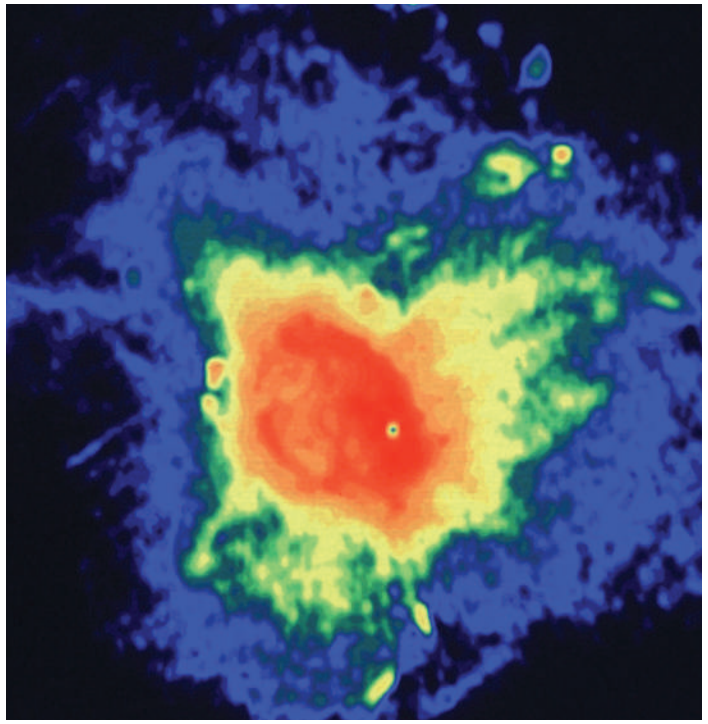
The source of these mammoth star clusters are giant molecular clouds — cool, dense complexes of dust and hydrogen gas up to 130 light years in breadth and containing the mass of between 10,000 and 500,000 Suns. Instruments such as those on NASA's Spitzer space telescope can see the parts that contain young stars, which glow brightly in the infrared. Astronomers working at longer wavelengths can see even earlier stages of the star-birth process. New maps of the Galactic Centre made at the Caltech Submillimeter Observatory (CSO) on Mauna Kea, Hawaii, reveal objects so early on in their development that they can't yet be called stars. "What we see are usually cores of clouds that have not necessarily begun to form stars yet, or are in the early stages of doing so," says Elisabeth Mills, who is a member of CSO's Bolocam Galactic Plane Survey. "With millimetre-wave data you get a more unbiased census of where star formation begins to occur. You see all of these nurseries, and whether or not they have a 'baby' in them yet."

A mosaic of these maps reveals a different side to the Galactic Centre. The bright clusters seen in Spitzer images are diminished; the cores of cooler clouds blossom with light, indicating that they have yet to collapse into massive protostars. These regions, which form a ridge-like structure in images taken at millimetre wavelengths, might one day form a chain of clusters like those in the Arches or Quintuplet, unless gravitational forces from the black hole disrupt their formation. Such structures, says John Bally, the principal investigator of the survey, "are unique to the Galactic Centre region."

The Galaxy's dark heart

The one structure that is absolutely unique, though, is the supermassive black hole itself. Given its apparent size and proximity to Earth, says Geoffrey Bower of the University of California, Berkeley, it affords astronomers their best chance to image the black hole's event horizon — the boundary beyond which no light can escape.

At the moment, the best observations of the black hole and its accretion disk are those made by a technique that links radio telescopes around the world, called very long baseline interferometry. At a distance of 26,000 light years, an interferometer working at radio wavelengths with a baseline the size of a planet should be able to resolve details as small as the orbit of the Earth. Unfortunately, the radio waves from Sgr A* pass through intervening regions of highly ionized gas, which scatter its radio emissions. "These random distortions blur the image of Sgr A*, much like frosted glass blurs an image," says Bower.



NRAO/AUI

At radio wavelengths, shells of gas surround Sgr A* (point source at centre).

Bower and his peers hope that new generations of interferometers working at millimetre and submillimetre wavelengths, which are less subject to intervening distortion, might in the long run actually reveal the black hole's event horizon[8]. How would that look to outside observers? Depending on its orientation, astronomers think that the relativistic effects of the black hole's intense gravitational field would make the event horizon appear as a large shadow or a silhouette cast on a background of bright plasma, in which the shadow is the boundary where light passes into the throat of the black hole itself.

"This image can be made with a network of millimetre- and submillimetre-wavelength telescopes distributed around the Earth," Bower says. Some of these telescopes already exist. Others, including the largest, the Atacama Large Millimeter Array, are under construction. Lashing them together into an ad-hoc interferometer the size of Earth, though, is a daunting technological challenge. That, says Bower, "is part of the thrill of the chase".

Such an observation would, of course, be a milestone — the first direct proof that an event horizon, and therefore a black hole, exists. And observing 'hot spots' orbiting the black hole would allow astronomers a qualitative way to test the effects of relativity in a strong gravitational field, an endeavour that has so far yielded ambiguous results[9]. The Galactic Centre may no longer be the mystery it was in Shapley or Jansky's day, but the better known it is, the more remarkable it looks — and it promises to become even more remarkable before too long. ■

Jeff Kanipe is a science writer based in Maryland.

1. Jansky, K. *Proc. Inst. Rad. Eng.* **232,** 1158–1163 (1935).
2. Shapley, H. *Contributions from the Mount Wilson Solar Observatory* No. 152 (1918).
3. Becklin, E. E. & Neugebauer, G. *Astrophys. J.* **151,** 145–161 (1968).
4. Lynden-Bell, D. *Nature* **223,** 690–694 (1969).
5. Muno, M. P., Baganoff, F. K., Brandt, W. N., Park, S. & Morris, M. R. *Astrophys. J.* **656,** L69–72 (2007).
6. Ghez, A. M. *et al. Astrophys. J.* **586,** L127–132 (2003).
7. Muno, M. *et al. Astrophys. J.* **622,** L113–116 (2005).
8. Bower, G. C. *J. Phys. Conf. Ser.* **54,** 370–376 (2006).
9. Broderick, A. E. & Loeb, A. *J. Phys. Conf. Ser.* **54,** 448–455 (2006).

Endangered collections

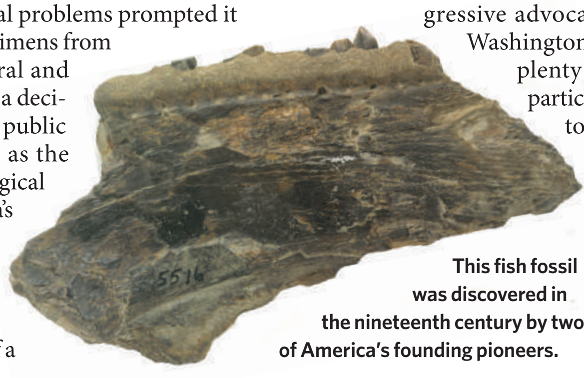
Philadelphia's venerable natural history museum is teetering on the brink of financial disaster. A new president recently took the helm, but can he save one of America's great institutions? **Rex Dalton** reports.

ACAD. NAT. SCI./ E. S. STEWART LIB.

The Academy of Natural Sciences in Philadelphia is best known for its vast natural history collections from some of America's founding pioneers: Benjamin Franklin, Thomas Jefferson, Meriwether Lewis and William Clark. But a visitor today might meet some decidedly less august figures coming through the door: express packages full of snails and slugs, sent to the museum regularly from pest inspectors at US borders.

When North America's first natural history museum was founded in 1812, no one envisaged that it would one day end up fighting invasive species. But the centuries have shown the huge importance of the academy's collection of 17 million specimens. Academy scientists rely on its vast holdings and library to identify animals and plants that have crept into new areas, and their expeditions have extended into far-flung corners of the world. Last year, for instance, a researcher from the academy reported finding the Tiktaalik fossil — a veritable missing link in the evolutionary journey from water to land some 375 million years ago — on a near inaccessible island off Greenland[1].

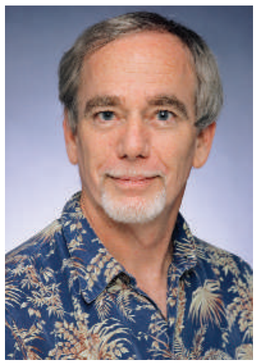
Yet despite the academy's rich history and modern discoveries, it is struggling to survive. Last year, financial problems prompted it to sell 18,200 specimens from its historic mineral and gem collection — a decision that became public at the same time as the city hosted Geological Society of America's annual meeting. The move infuriated scientists, many of whom saw it as selling off a precious piece of history.



This fish fossil was discovered in the nineteenth century by two of America's founding pioneers.

ACAD. NAT. SCI./ E. S. STEWART LIB.

Now, though, the sale seems to have marked the academy's low point. A new president, William Brown, arrived at the helm in January, bringing with him a strong background of having righted a previously troubled museum. For the past five years Brown, an ornithologist with a law degree from Harvard, served as president of the Bishop Museum in Honolulu, the official state museum of Hawaii. During his term, Brown extricated the museum from its financial and public-relations challenges. "He really left the museum in a better place," says Haunani Apoliona, a Bishop board member who is also an elected trustee of the state's Office of Hawaiian Affairs. "Everyone is sorry to see him move on."



William Brown has a history of rescuing museums.

BISHOP MUSEUM

Brown seems to have his work cut out for him in Philadelphia. "The academy has been waiting for someone like Bill for a long time," says museum trustee Thomas Lovejoy, a biologist and president of the Heinz Center, a progressive advocacy institute in Washington DC. "There is plenty for him to do, particularly attending to a bruised staff that has been through some rough years."

Financial woes are the first of many problems. At about US$55 million, some say that the academy's endowment is lower than it needs to be to support a research institution of its size. For the past decade, academy deficits have averaged some $700,000 a year, although it did end up $2 million in the black for 2005 on its $17-million operating budget. The debts have taken their toll on staff; over the past ten years, curators have been laid off in waves, with three of the ten remaining being laid off in 2005.

The mineral sale netted the academy more than $1 million, which will be used to bolster its renowned library. But with Brown now in place, a proposed sale of another 7,300 specimens has been halted. This collection of North American minerals was donated by the Vaux family nearly 125 years ago, who stipulated that they must never be sold. That kept the collection out of the October sale, and now Brown and academy scientists are drawing up plans to exhibit the minerals for the first time in decades. Many academy scientists are relieved; selling the Vaux collection on top of the other minerals, says palaeontologist Ted Daeschler, would have been "too horrible to swallow".

In private, many academy scientists lay part of the blame for the museum's desperate financial straits on its board of trustees — a typical mix of local business executives, community leaders and a few scientists. Some academy staff say that board members were well-meaning but sometimes uninformed, which hindered their ability to engage fully in the museum's problems.

Academy staff have also faced tough times with the museum's presidents. Before Brown's

arrival, administrators sometimes threatened to fire anyone who spoke out publicly about the museum's problems. Brown's predecessor, former US oceans chief James Baker, left in April 2006 after serving for four years. Administrator Ian Davison then stepped in to act as chief executive temporarily, overseeing the mineral sale among other changes.

Troubled times

The academy is not the only museum facing tough times (see page 594). In recent years, natural history museums in the United States have struggled to survive in the face of diminishing government support, competition from the entertainment industry, and lack-lustre philanthropy. Some have taken what scientists deridingly call the theme-park approach — glitzy exhibits to pull in paying crowds, while the scientific mission is reduced if not all but eliminated.

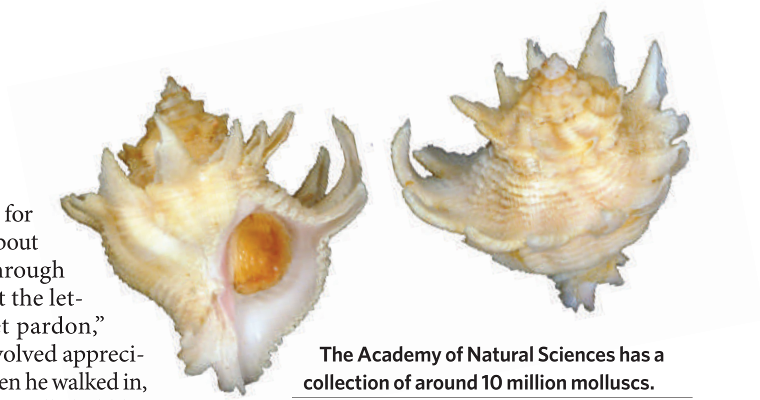
Bitter root (*Lewisia rediviva*) was named after Meriwether Lewis.

Museums have also used varying approaches to pay back bonds issued for refurbishment, such as increasing their revenue with community dinners or party events, and bringing in star-quality exhibits to boost attendance. For instance, the Houston Museum of Natural Science is taking the controversial step of creating a touring exhibit that includes the real skeleton of Lucy, the famed 3.2-million-year-old hominid from Ethiopia[2].

Brown comes to Philadelphia with a reputation as a savvy financial manager. At the Bishop Museum, he accelerated programmes to rent out museum locations for parties. "We went from $50,000 to a $1 million a year in revenue," he says. And although Philadelphia isn't exactly sunny Honolulu, the downtown part of the city is undergoing a major redevelopment, offering new revenue opportunities from businesses and visitors.

Brown also comes with a track record of smoothing over museum disputes. When he took on the position at Bishop, the museum was embroiled in its own volatile controversy — about 80 priceless Hawaiian artefacts had been illegally spirited away and hidden in an isolated cave by indigenous traditionalists. By the time Brown departed, he had managed to help secure the return of the specimens, after a lengthy legal battle.

Brown's predecessor at the academy had sanctioned numerous employees with letters in their personnel files for speaking out publicly about the dispute. "I went through all the files, pulled out the letters and did a blanket pardon," says Brown. Those involved appreciated his openness. "When he walked in, the artefact dispute was really bubbling over," says Apoliona. "His efforts were really honourable. Everything was transparent."



The Academy of Natural Sciences has a collection of around 10 million molluscs.

In Philadelphia, Brown will face the challenge of maintaining the remaining staff. Some of those who haven't been fired have left of their own accord, such as renowned botanist Lucinda McDade. She left the academy last year to be research director at the Rancho Santa Ana Botanic Garden in Claremont, California. "I really missed the teaching," she says. In Philadelphia, almost no graduate-education programmes are associated with the academy, even though it has plenty of top universities nearby, such as the University of Pennsylvania.

Brown says that he is interested in exploring the possibility of bolstering graduate programmes. "The academy has many positive aspects," he adds. "It is clear we need to get all the cylinders firing effectively." He knows the academy's legacy well; he studied there briefly as a young scientist, was on the board between 1987 and 1993, and remained an adviser in recent years. His experience with government funding agencies will also come in handy — he was science adviser to Bruce Babbitt, secretary of the US Department of the Interior under President Bill Clinton.

Brown will have a portfolio of top scientists to handle. In addition to Daeschler, who discovered the Tiktaalik, academy ichthyologist John Lundberg is completing a family tree of the catfish, a creature that can be found on every continent[3]. He, together with the academy's Mark Sabaj and their colleagues, regularly plies the waters of the Americas, Africa and Asia, confirming previous sightings and discovering new species. The team expects to identify 1,750 new species, which will push the total number of catfish species up to 4,600.

Pink fairies (*Clarkia pulchella*) were named after William Clark.

Illegal immigrants

Elsewhere at the academy, the malacology collection — of snails, slugs, clams and various other molluscs — includes 10 million specimens, some of which date to the 1700s. These specimens, along with the library, help malacologist David Robinson to guard ports of entry for the US Department of Agriculture. It is his express packages that arrive at the academy, at a rate of up to 60 a day, full of invasive species. The crawlers and sliders try to slip in across the border in shipments of marble, tiles, stones, manufactured goods, produce, plants, and even military vehicles — arriving from all continents. When the border inspectors can't identify the species, they send them to specialists such as Robinson for help.

Because he spends so much of his time identifying unknown critters, Robinson rarely publishes on the rare and unusual species he discovers — like one in 1999. That express package contained dried plants collected by scientists from the Missouri Botanical Garden in St Louis and flown from Bali to the port of entry near San Francisco. Buried in the leafage was a lone dead snail. Robinson was stumped: "I didn't have a clue what it was for months."

Then, leafing through a publication[4] one day, he spotted the Bali native (*Anaglyphula whitteni vermeulen*) and was finally able to identify his specimen. And Robinson's work on that snail is but one example of how a long-standing academy can help scientists in the new century to fight invading pests. ■

Rex Dalton is *Nature*'s West Coast correspondent.

1. Daeschler, E. B., Shubin, N. H. & Jenkins, F. A. *Nature* **440,** 757–763 (2006).
2. Dalton, R. *Nature* **444,** 8 (2006).
3. Lundberg, J. G. *Neotrop. Ichthyol.* **3,** 597–605 (2005).
4. Vermeulen, J. J. *Basteria* **59,** 149–162 (1996).

NATURE|Vol 446|5 April 2007

M. KLUM/NATIONAL GEOGRAPHIC/GETTY

Logging: the new conservation

Can a vast monoculture plantation be at the forefront of biodiversity protection? **David Cyranoski** meets conservation biologists who hope to save species by making peace with the enemy.

The forest in Borneo is usually a lush affair. But in the province of Sarawak, some forest regions have a striking monotony. Here, the skinny trees are all the same variety of acacia, some 20 metres high with short, jutting branches. The trees are spaced in regular rows, three metres apart.

Acacia mangium, a fast-growing species that is good for pulp and paper, is not native here. But at these nurseries on the Malaysian part of Borneo, 70,000 hectares of forest have been razed and replaced with a monoculture of acacias, grown at a rate of 3 million seedlings a month. Grand Perfect, a consortium owned jointly by three local timber companies, eventually plans to cover a few hundreds of thousands of hectares like this. The acacias will grow for just seven years before being harvested. Eventually, 30-tonne trucks will come every few minutes to carry away 3.5 million tonnes of wood a year.

This factory-like image is not what an environmentalist might call conservation. But many scientists are calling it the future of biodiversity protection. It is Grand Perfect's Planted Forests Project, a 490,000-hectare experiment in which just under half the land is set aside for logging, around a third for conservation and the rest for use by indigenous groups. In this unusual mosaic, customary foes — timber companies, conservation biologists and local tribes — are working together to create a forest-management scheme that can meet all their respective needs. Each group has much to gain, or lose. For conservationists, the project's land is a precious piece of southeast Asia's famously biodiverse tropical forest, an important 'hotspot'. Globally, tropical rainforests are disappearing at a rate of 12 million hectares a year. The Sarawak government, which commissioned and funds the project, and the timber companies that it has contracted, stand to gain tens of millions of dollars, if not more. Those living in the forest want to preserve their right to roam, hunt, fish and farm, and are being promised infrastructure and employment.

Rob Stuebing, a herpetologist who is the head of Grand Perfect's conservation division, is in the uncomfortable position of trying to reconcile these demands. It requires careful diplomacy. He argues with the loggers and pleads with the locals. He also deals with conservationists and fellow scientists, some of whom share his excitement about this new approach. Others charge that the effort is just an attempt to put an environmentally friendly gloss on a typical logging project.

"You can't change a whole system by working against it. The only way to do it is from the inside." — Rob Stuebing

D. CYRANOSKI

Stuebing's greatest challenge will be to prove the sceptics wrong. He needs to show that biodiversity can be preserved within an area devoted mostly to a non-native monoculture, something that has never been proved before (see 'Tree swap'). As a conservation biologist taking up a position with a timber company, is Stuebing shaking hands with the devil? "I had misgivings at first," says Stuebing. "But you can't change a whole system by working against it. The only way to do it is from the inside."

Stuebing is a 30-year veteran of Malaysia's natural history who also has experience in the timber industry. He is tall, lean and seems more comfortable dressed for the field than in the suit and tie that his current job sometimes demands. He also has little patience for the most idealistic of his critics, who say that true conservation would be to enforce a hands-off policy. Since he joined Grand Per-

fect in October 2004, Stuebing has moved full-speed ahead on his mission: an inventory of the wildlife in the massive planted forest zone and coordination with the timber companies on a foresting strategy that protects, however imperfectly, what he finds. The plantations were started in 1996 without any biodiversity project in place, and some critics have said that the inventory is starting tens of thousands of hectares too late.

In January, Stuebing held a conference called Biodiversity Conservation in Tropical Planted Forest in Southeast Asia. Held in Bintulu, a coastal town 45 kilometres from the inland planted forest zone, the meeting was the first opportunity for all interested parties to discuss the future directions of his project and work out how to assess its progress. Some 70 naturalists, taxonomists and conservation biologists as well as representatives from the timber companies and the government attended the conference. The Sarawak chief minister, Pehin Sri Haji Abdul Taib Mahmud, called it a "dream come true for Sarawak's forest and biodiversity conservation" and, in grand ceremonial fashion, planted the plantation's 88 millionth tree. The local authorities love the scheme because they can make money from the land and still feel good about conservation.

R. STEUBING

The Planted Forests Project will mingle pockets of acacia trees (above) with existing tropical forest.

Chop and change

Like many lush tropical forests, much of the area allocated to the planted forest zone is already 'disturbed'. For centuries, indigenous groups have chopped up parts of the forest to make room for farmland. Large-scale industrial logging, starting some 40 years ago, took a greater toll. Original 'old-growth' forest occupies at most 20% of the area; what has grown back on the disturbed land is called secondary forest.

The idea that these forests are already degraded and will continue to change is an important, and controversial, one. Some experts say that trying to return biodiversity to some ideal past era is unrealistic. Jianguo Wu, a specialist on sustainability at Arizona State University in Tempe, opened the meeting with an appeal to consider forests as something to be used for sustainable development rather than as a static entity. Others vehemently disagree, and say that in the past five or ten years the degradation has become a frequently used excuse to convert forests to plantations in Malaysia and Indonesia. "All the forests need is some rest," said one silviculturist who did not want to be named.

Everyone agrees that there are limits to what can be called biodiverse. On the other side of the highway from one of the acacia plantations lies a wasteland of terraced, orangeish soil — the scar of an oil-palm plantation. Malaysia is the number one producer of oil palm, but this type of forestry is the worst-case scenario for the conservationists. While Hollywood stars talk up the environmental friendliness of biofuels such as palm oil, here, the plants that produce them are demonized because they leave little room for anything else to live. If nothing else, everyone hopes that the acacia plantations will be more forgiving to their environment.

But what is at stake? Borneo is home to some 1,300 known mammals, birds, frogs, snakes and lizards, about two-thirds of which are thought to be in and around the planted forest zone. Stuebing estimates that around 20% of the mammals and perhaps 60% of the amphibians are endemic, existing here and nowhere else. But these are approximations. The area has never been explored thoroughly, and the project's first priority is to take an inventory. As the loggers move in, there is a sense of urgency because many species could be lost before they are ever discovered.

> "Some romantics, like myself, thought that all was lost with the cutting of the rainforests."
> — Robert Inger

At the conference, Stuebing called on all scientists to help identify the forest's wealth of organisms. Many, keen to expand their collections, have already taken up the call and their initial forays have turned up plenty. Jaap Vermeulen of the National Herbarium of the Netherlands at Leiden University has found 83 species of snail in a region called Bukit Sarang, part of the area devoted to conservation. "Thirty-one of these are new to science. It is exceptional, a really hot piece of real estate," he says.

Stuebing's challenge is to use the conservation zones, which are marbled into the plantations, to simulate a more extensive forest that has room for large mammals to roam. When local populations go extinct, larger forests also support 'metapopulations' that can repopulate them. Stuebing works with the loggers to designate corridors — such as 50 to 500-metre-wide swathes of land alongside rivers — that are crucial for mobility of many species. Maintaining the survival of rare, symbolic flagship species, such as hornbills, tufted ground squirrels and gibbons, would be a huge success. But Stuebing is not sure that his corridors will be long or wide enough.

Support structures

The biggest question mark, and one that could be crucial to the success of the entire venture, is whether the acacias will support wildlife. Frogs, rats, squirrels, ungulates and carnivores seem to do well in acacias because they like the food and breeding sites. Understory birds, snakes, lizards and bats, however, are less happy. Bill McShea, an ecologist at the Conservation and Research Center of the Smithsonian National Zoological Park in Front Royal, Virginia, uses camera trapping to follow large mammals. He found that about two-thirds of the 39 species he had identified in the local 'disturbed' forest

N. LUCAS/NATUREPL.COM

Cutting down: large areas of primary forest are being cleared to make way for acacia plantations.

were also living in the acacias. "So does [this result] mean they are running for their lives? Or are they living there stably? It is impossible to tell right now," he says. And even if the animals are settled, no one knows what will happen when the acacias are harvested after seven years.

Stuebing admits that, with only 2% of Grand Perfect's total budget, the conservation programme is a relatively small concern for the timber companies, so working to influence them is crucial. But the companies are coming under increasing pressure from their buyers to certify that pulp and paper exports come from rain forests that are managed sustainably. Such concerns give Stuebing more leverage to fight for wider corridors. Grand Perfect says that it is also trying to change its logging practices by, for example, using an elevated cable system to extract logs rather than dragging them on the ground and eroding the soil.

Lap of luxury

Grand Perfect's efforts to integrate scientists — and its construction of relatively luxurious field stations — impressed the conservation biologists who attended the Bintulu meeting. "Never before has a commercial enterprise spent so much time ensuring that there is a biodiversity programme in place," says the Earl of Cranbrook, chairman of the International Trust for Zoological Nomenclature and a 50-year veteran of the southeast Asian natural-history scene.

Because the acacia plantations are very productive, some argue that they will also relieve the pressure to cut into other, more pristine forest. Warren Ellis, who represented Grand Perfect's logging division, told the meeting that a hectare of acacia plantation produced more wood than ten hectares of forest that had been logged and had then regrown naturally. Hans ter Steege, a plant ecologist at the National Herbarium of the Netherlands in Utrecht, is also optimistic. "I would approach the project with a positive attitude. All the right ingredients — social, economic and ecological — have been addressed."

But not everyone is happy. Some scientists have accused the project of failing to take into account a key factor — the incredible diversity of the indigenous trees. Peter Ashton, a retired expert on the forests in Sarawak, points out that the Sarawak lowlands contain some 3,000 species of tree and that the trees support the insects and microorganisms that account for the vast majority of the forest's biodiversity.

Token gestures

Conservationists have also become cynical of nods to their cause by loggers. A 2006 report by the Center for International Forestry Research in Indonesia found that some companies are recruiting billions of dollars from investors by saying that they have sustainable sources to feed their pulp mills when, in reality, they are relying on natural forests in Indonesia for most of the supply. Gustavo Fonseca, the chief conservation and science officer at Conservation International in Arlington, Virginia, questions whether the proportion of land given over to native forest in the project is high enough, and says that road-building and increased fire-hazards "may create impacts far beyond [those of the] plantation". Fonseca also notes that *A. mangium*, although not thought to be invasive, has shown signs of spreading spontaneously in the Caribbean.

As yet, there is no pulp mill, and the logging will probably not start in earnest for several years. By then, Stuebing and his colleagues will have a better idea of whether their patchwork experiment will work. If it does not, the acacia harvesting will still proceed. The data obtained could still be of use, because they will probably be the best record yet of how acacia plantations affect biodiversity, something that is important for the many other plantations in Sarawak and elsewhere in Borneo. And Stuebing looks on the bright side: "At the very least we have done some good science."

The conservationists may not share his rosy outlook, but they — and the loggers they traditionally eyed with aversion — may also gain a new perspective. "Some romantics, like myself, thought that all was lost with the cutting of the rainforests," says Robert Inger of the Field Museum in Chicago, who has worked in Borneo for decades. Like most at the conference, Inger now recognizes that demand for timber means the inevitable end of many species. He has watched frogs try to familiarize themselves with their new environment in the acacias and is hopeful that they have set up shop there. "We may have to look at the forests again with less romantic eyes," he says. ■

David Cyranoski is *Nature*'s Asia-Pacific correspondent.

See Editorial, page 583.

Tree swap

Plantations in biodiversity zones do not usually take pains over conservation. There are a few examples in which survival of native species has been measured, but their results have been discouraging.

The Jari project in Brazil is one of few precedents on the same scale as Malaysia's Planted Forests Project. In the 1960s, US billionaire Daniel Ludwig purchased 1.7 million hectares of almost undisturbed forest and used about 10% of it to plant trees for paper. Ludwig chose *Gmelina arborea* — a non-native, fast-growing tree — but it turned out that it did not grow well in the Amazon and was replaced with eucalyptus.

At the time, the project did not have a biodiversity programme in place, but in 2003, Carlos Peres, an expert on tropical forest ecology at the University of East Anglia in Norwich, UK, launched one with the Emílio Goeldi Museum in Belem, Brazil. Peres's group looked at more than 1,450 species, but with disappointing results. "Suffice it to say that, for the 16 taxonomic groups of plants, vertebrates and invertebrates that we sampled, primary forests were irreplaceable in terms of the proportion of species that can be retained," he says.

On the Tiwi islands in Australia's Northern Territory, native eucalyptus has been replaced with *Acacia mangium* on a much smaller scale than that of the Planted Forests Project.

But the plantation has displaced much of the native wildlife. "The environmental consequences have been profound," says John Woinarski, from Australia's Department of Natural Resources, Environment and the Arts, and an author of a study into the project's effects (R. Firth *et al. J. Biogeogr.* **33,** 1820–1837; 2006). In particular, most of the mammals have made little use of the plantation, including species such as the brush-tailed rabbit-rat and Butler's dunnart, which are considered threatened or vulnerable. **D.C.**

Follow Thompson's map to turn biology from a science into a Science

SIR — Evelyn Fox Keller, in her Essay "A clash of two cultures" (*Nature* **445,** 603; 2007), argues that biology may not have general laws and that the approach that is natural to physics would probably not work in biology.

In the opening of his seminal 1917 book *On Growth and Form*, D'Arcy Thompson quoted the eighteenth-century philosopher Immanuel Kant, who lamented that the field of chemistry had not yet embraced a mechanistic and mathematical expression of chemical phenomena. As a result, according to Kant, chemistry at that time was just a science, rather than a Science with a capital S. Despite Kant's view, however, as Thompson emphasizes, a great quantitative revolution proceeded to transform chemistry into a capital-S Science every bit as rigorous as physics. Thompson goes on to argue that biology is poised for just such a quantitative revolution.

Today, Thompson's thesis is being borne out; biology is becoming an increasingly rigorous quantitative Science that is finding more generality with each publication cycle. Most would agree that mathematical theories of quantitative genetics (including the modern synthesis), populations dynamics, organism interactions, epidemiology, ecosystem processes and growth and metabolism have together revolutionized biology, transforming it into a capital-S Science. This quantitative revolution would have been greatly muted, though, had investigators not been compelled, by Thompson's explicit advice, to identify general patterns and laws, to describe these quantitatively and to search for underlying mechanisms.

In this light, Keller's thesis that biology is a series of exceptional cases is a great leap backwards.

Instead of following Keller's philosophy, biology needs to adhere to Thompson's original roadmap and continue its transformation into a rigorous and quantitative Science. It is not fundamentally different from the physical sciences. Keller is correct to note that biology is often unique, but historical contingency does not preclude the existence of general patterns or mechanisms, and has not limited the development of the arguably more rigorous and quantitative fields of astronomy, geology and economics.

Before accepting Keller's thesis, one must first reject the alternative hypothesis that there are general patterns and that these patterns have their basis in equally general mechanisms. For example, Keller dismisses metabolic networks as one possible law-like phenomenom because 'power laws' are not as ubiquitous as first thought. This is conjecture: a quick glance at the biological scaling literature associated with organismal metabolism, life-history, ecology and even evolutionary dynamics reveals an impressive series of general power-law-like behaviours apparently interrelated by a common mechanistic framework based on organismal metabolism. If there are general patterns with equally general mechanisms, then arguably biology has laws. With bioinformatics, large-scale research networks and new computational techniques, we can rapidly identify the existence of general patterns and illuminate the mechanisms that underlie them.

Reasons to follow Thompson's roadmap for biology could not be more urgent. The need to understand and predict the response of the biosphere to climate change, the spread of emerging diseases, the collapse of biological diversity and the need to improve the human condition through medicine and agriculture — all these demand the development of a quantitative, mechanistic and predictive biology. The viewpoint espoused by Keller does not advance the development of biology into a Science. The future of biology belongs to those who follow Thompson's roadmap.

Brian J. Enquist, Scott C. Stark
Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, Arizona 85721, USA

Authors defend study that shows high Iraqi death toll

SIR — In our opinion, your News story about our *Lancet* paper "Death toll in Iraq: survey team takes on new critics" (*Nature* **446,** 6–7; 2007) has confused the matter rather than clarified it. You outline three criticisms of our work: that there was not enough time to have conducted the survey; that the sampling method suffered from a 'main-street bias'; and that the study team fabricated the data (the last being attributed to anonymous "researchers"). These criticisms have been previously addressed, and have little merit.

On the first point, the 1,849 interviews in 49 days described in our study suggest that 38 interviews had to be conducted each day by our eight interviewers. Although introducing themselves and explaining the confidentiality agreement might have taken interviewers several minutes, the five-question interview would take only a couple of minutes for most households that reported no deaths. The idea that eight interviewers could not conduct a total of 38 interviews in a day is not credible.

Second, we dismiss the suggestion that our sampling over-represented main streets, where car bombs are more likely. As stated in our paper (G. Burnham, R. Lafta, S. Doocy and L. Roberts *Lancet* **368,** 1421–1428; 2006), when excluding the statistically outlying cluster of Falluja from the first report, we estimated 98,000 (95% c.i.: 8,000–194,000) excess deaths versus 112,000 (95% c.i.: 69,000–155,000) over the same period with the second survey. The first survey was done selecting random starting points with a Global Positioning System unit. The second used the random street-selection process, which is being criticized as biased. It rarely occurs in the field that two sampling methods are used allowing for comparison, and here the results are nearly identical. Moreover, there is no plausible mechanism for a significant main-street bias to operate, because only 15% of all deaths are from car bombs and other ordnance, and because most violent deaths are believed to occur away from the home.

Third, as for the accusation that researchers fabricated the data, we are ready, willing and eager to have an established international authority take a sample of the cluster forms and go to the field with our interviewers to verify the findings. Until that time, the Coalition and Iraqi governments' statements that during the first three years of occupation, Iraq's violent-death rate was lower than those of Russia, Estonia, Latvia, South Africa and Kazakhstan remain an implausible contrast with our findings.

When *Nature* called one of our study members in Iraq and asked if local officials joined them during the survey, that individual later clarified to *Nature* by e-mail that 'local officials' did not mean local clinicians and colleagues. This was inaccurately reported in the *Nature* summary along with a statement by our co-author that interviewers often worked alone. These points were wrongly cited as contradictions between the study team members in your News story.

All reports will eventually have "criticisms that dogged the study", if previously addressed criticisms with so little merit are given a voice in the press.

Les Roberts*, Gilbert Burnham†
*Program on Forced Migration and Health, Mailman School of Public Health, Columbia University, New York, New York 10032, USA
†Center for Refugee and Disaster Response, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland 21205, USA

***Nature* stands by its version of the events described in the penultimate paragraph of the above Correspondence — Editor, *Nature*.**

Contributions to Correspondence may be submitted to correspondence@nature.com. They should be no longer than 500 words, and ideally shorter.

BOOKS & ARTS

C. TUTTLE/CORBIS

Clear benefit: maintaining a clean environment should be a political priority.

Avoiding system failure

An upgraded version of capitalism is needed to protect the world's resources.

Capitalism 3.0: A Guide to Reclaiming the Commons
by Peter Barnes
Berrett-Koehler: 2006. 216 pp. $22.95

Robert Costanza

An astronomical number of words has been written on the complex set of problems facing humanity, such as climate change, peak oil, poverty and AIDS, especially in comparison with the number devoted to serious solutions. There is also a growing recognition that these problems are systemic — solving them will require changes to the way society operates, the basic capitalist system.

Peter Barnes is a progressive businessman who created a successful company, Working Assets, that donates 1% of its revenue (not just its profits) to environmental causes. Now he has produced a landmark book that gets to the heart of one of the most important systemic problems of the current capitalist system, and proposes workable solutions. He argues that the previous (1.0) and current (2.0) versions of capitalism that evolved up to the twentieth century, under conditions that no longer hold, are in serious need of an operating-system upgrade. Version 3.0 must address the conditions and problems we face now. Barnes fully recognizes the benefits of capitalism, and does not recommend replacing it wholesale with something completely different. But he also recognizes the major flaws in version 2.0, and describes some of the features that will be needed in version 3.0 to fix them.

A major flaw, according to Barnes, is the failure of the current system of capitalism to adequately value and manage the commons. The commons, or more descriptively the common wealth, includes the atmosphere, clean water, oceans, airwaves, social networks, cultures and many other things that are essential to sustainable human well-being, but that either cannot or should not be put under private ownership (in other words, our natural and social capital assets). The failure of communism can be traced to the desire to put too many assets into the commons. But capitalism 2.0 puts too few of humanity's assets into the commons, or leaves them as open access to be plundered by private interests. Barnes' capitalism 3.0 would recognize the importance of the commons as a separate and distinct sector of the economy that deserves to be valued and 'propertized' — but not privatized. The commons can be treated as property without privatizing them by creating various kinds of common property institutions, including trusts.

An analogy might help at this point. What if your computer's operating system started inadvertently overwriting some of the system files that allowed it to function? At first, this may not be a big problem, or even noticeable. But eventually it will cause the system to crash. This is Barnes' analogy for the operation of the current capitalist system. By inadvertently depleting our natural and social capital assets, which are external to the market and therefore unvalued by the market, capitalism 2.0 is leading us towards a system crash. We need to upgrade the system to stop and reverse this depletion and prevent the crash.

One could argue that governments (especially democratic ones) already represent the public interest and the commons sector of the

economy, and in some cases may even over-represent it. But governments have too often become the representatives of private corporate (rather than public) interests and are thus not performing this function adequately, leaving the commons underrepresented, undervalued and underprotected.

Barnes' basic recommendation is the creation of trusts to manage common property rights. An example is the Alaska permanent fund, a trust set up by the state of Alaska to manage royalty payments for oil and gas extraction. Other examples include the various land trusts, easements and concessions that have evolved and been used effectively by non-governmental organizations such as the Nature Conservancy and Conservation International. Barnes' vision would greatly expand these modest beginnings to cover all our common natural and social capital assets (while leaving private property intact). These common assets are estimated to be worth several times more than all the world's private assets combined (see *Nature* **387,** 253–260; 1997).

A good example would be an atmospheric trust to manage carbon dioxide (and other air pollution) emissions. The various cap-and-trade systems proposed so far for CO_2 would fit nicely into the trust scheme, especially if permits were auctioned to emitters, rather than given away free of charge. The trust would set the caps (reducing them over time), collect the auction fees (which would increase as the cap was reduced), and use the revenues to protect and enhance the asset (by investing in renewable energy, for example, or paying private landowners to sequester carbon). The trust could also distribute a fraction of the revenues to all citizens on a per capita basis, as Alaska does now with oil and gas royalties, in the form of tax credits and dividends. This would reverse the regressive nature of fees on carbon emissions because large emitters would pay more into the trust than they got back, whereas small emitters would get back much more than they pay in. On the global scale, this payment could also help to substantially alleviate poverty.

The details of how to pay individuals while avoiding bureaucratic corruption and excessive transaction costs would need to be worked out, of course, and this would be no mean feat. But this is just the kind of challenge that creative capitalists (version 3.0) would warm to.

Upgrading capitalism will not be without trauma (think of the last time you upgraded your computer system), but it is long overdue and is essential if the capitalist system is to survive another century. There are many devilish details to work out, but the upgrade is already under way and gaining momentum. Let's hope the system doesn't crash before it can be completed. ■

Robert Costanza is director of the Gund Institute for Ecological Economics, Rubenstein School of Environment and Natural Resources, University of Vermont, 617 Main Street, Burlington, Vermont 05405-1708, USA.

Neurons and knowledge

Second Nature: Brain Science and Human Knowledge
by Gerald M. Edelman
Yale University Press: 2007. 203 pp. £16.99, $24

David Papineau

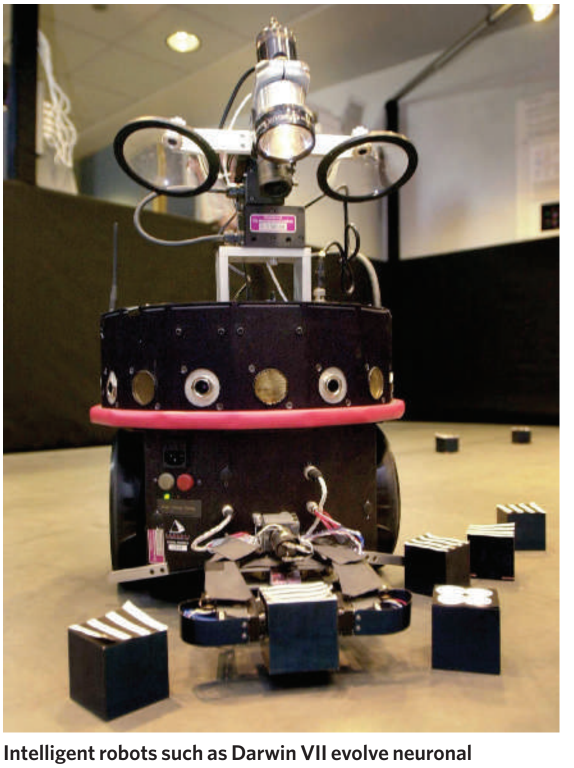
In 1972, at the age of 43, Gerald Edelman won a Nobel Prize for his work on the molecular structure of antibodies. Since then he has written a series of books about the human mind, starting with *Neural Darwinism* (Basic Books, 1987) and including *The Remembered Present* (Basic Books, 1990), *Bright Air, Brilliant Fire* (Basic Books, 1992) and *Wider than the Sky* (Yale University Press, 2004). In his latest book, *Second Nature*, Edelman turns his attention to epistemology. His aim is "to ground the theory of knowledge in an understanding of how the brain works".

Epistemology, or the theory of knowledge, is traditionally the preserve of philosophers, and is a house with many mansions. At the start of modern philosophy, René Descartes aimed to show how our beliefs could be placed on absolutely firm foundations. More recently, philosophers have become suspicious of Descartes' aspirations to certainty. As a result, many now agonize about how we should define 'knowledge' if it doesn't require certainty. Others seek to develop epistemology as part of probability theory, analysing the relation between evidence and estimates of the likelihood of events. And there are others still who see epistemology as an essentially practical enterprise whose task is to sift accurate modes of thought from unreliable ones.

Edelman sees little value in any of these philosophical efforts. He dismisses most of them as "armchair operations" that are not empirically testable. Even the more practically minded philosophers are criticized for ignoring the 'neural underpinnings' of knowledge. In Edelman's view, only a 'brain-based epistemology' will yield real results. In principle this sounds like an exciting project, but in the end the book is disappointing. Edelman covers a lot of ground in relatively few pages, but it is not always clear what he is trying to say.

The first few chapters focus on consciousness, on the not unreasonable grounds that knowledge presupposes consciousness. As will be familiar to readers of his previous books, Edelman takes consciousness to depend on 're-entrant activity' resulting from reciprocal signalling between different regions of the brain. It is hard to know what to make of this theory. Unsympathetic critics will ask what distinguishes re-entrant activity from the feedback projections that are familiar to all brain researchers. We might also ask how the theory measures up to Edelman's own high standards of scientific testability. For example, his theory implies that frogs are not conscious, as they lack sufficiently complex brains. But how are we to test this claim, in the absence of any independent criteria for frog consciousness?

When Edelman turns to knowledge itself, he is not much more explicit. Early on he promises "an account of knowledge that relates truth to opinion and belief, and thought to emotion by



NEUROSCIENCES INST., SAN DIEGO

Intelligent robots such as Darwin VII evolve neuronal connections in response to environmental stimuli.

including aspects of brain-based subjectivity in an analysis of human knowledge". It doesn't get much more definite than this. Edelman provides some general observations about neuronal development, pattern recognition and the evolution of language, but he doesn't explain exactly what epistemological questions these observations are supposed to answer.

The most substantial part of the book is a section on intelligent robots. Edelman describes a series of devices with sensors and computer-simulated brains that have been built in his Neurosciences Institute in La Jolla. The brains are modelled on human anatomy, complete with versions of visual cortex, inferotemporal cortex and hippocampus. They are not preprogrammed, but evolve neuronal connections

in response to experience. Edelman explains how these devices can learn to recognize patterns and navigate new environments. This is intriguing, but it is rather difficult to assess its significance in the absence of any comparison with other learning-based robotics research.

This lack of scientific comparisons is a striking feature of the book. A naive reader could easily form the impression that Edelman and his associates are the only people trying to use scientific information to cast light on the human mind. Not that the book lacks references: it is full of brief digests of philosophical and cultural figures, such as Wilhelm Dilthey, Ludwig Wittgenstein, Isaiah Berlin, Richard Rorty and many more. It is impressive that Edelman should have read their work, but I couldn't help wondering who his own book was written for. Is there really a readership for potted cultural history written by senior scientists? In any case, this is not a book for those who want to know how brain science relates to epistemology. ■

David Papineau is professor of the philosophy of science, King's College London, London WC2R 2SL, UK. He is the author of *Thinking About Consciousness* (Oxford University Press, 2002).

FILM

Dark days ahead

Sunshine
directed by Danny Boyle, written by Alex Garland
Twentieth Century Fox. Worldwide release from 6 April 2007.

Richard Webb

"Our Sun is dying," intones a laconic voice-over at the start of the film *Sunshine*. Why, in the year 2057, Earth's star is failing is unspecified. Astrophysical orthodoxy would demand that life on Earth is extinguished by the Sun's red-giant phase long before its light fades away — and that only in billions of years' time. But as Brian Cox, a researcher at the particle-physics laboratory CERN who acted as scientific consultant to the film, was at pains to point out in the pre-release publicity, there is still a lot we don't know about the Universe. The cosmos, like the cinema, is a medium of infinite surprises.

Not that *Sunshine*, at least in its basic scenario, is surprising at all. This is cinematic *déjà vu*: with humanity's survival threatened, a group of brave, bright young things must take up arms for all our sakes. The eight-man crew of the spaceship Icarus II has been charged with restarting the Sun. They have to drop a nuclear payload — fissile material amounting to the mass of Manhattan — into its core. A first attempt, Icarus I, has failed, with all contact having been lost seven years before.

Early enough in this second, last-gasp mission, a law of cinema as ineluctable as the law of gravity kicks in: what goes up never looks like coming down. Icarus II's calamities begin as it prepares for the slingshot round Mercury that will propel it into the Sun's corona: it picks up a distress signal from Icarus I. Diverting course towards their lost predecessor engenders a chain of events that puts both mission and crew in mortal danger.

But like Icarus II, the film manages to diverge from its preordained trajectory. Despite its big-budget look and feel, this is a (relatively) cheap-and-cheerful British production, which frees it from some of the more egregious demands of Hollywood schmaltz. In particular, its commitment to a happy ending is equivocal; each setback to the mission also demonstrably reduces the crew's own chances of survival. "There will be no ticker-tape parade," says Mace, the hard-boiled engineer, with bitter resignation as a particularly cruel sideswipe of fate blows out most of the ship's oxygen. That fatalistic undercurrent makes for an interesting study of human interactions in a — literally and metaphorically — pressurized environment.

Science is handled pleasingly well in *Sunshine*. Scientific vocabulary is not flung about with the gay, context-free abandon that is traditional in such circumstances. Before shooting the film, Cillian Murphy, who plays physicist Capa, guardian of Icarus II's nuclear payload, visited CERN to observe real physicists in their natural habitat. The gratifying (if somewhat surprising) result is a scientist on the silver screen who seems manifestly sane. It is true that when I worked at CERN, none of my physicist colleagues there had quite such rippling biceps, nor did they pout in the face of adversity quite so provocatively as Capa. But this much can be forgiven in the exercise of creative licence.

As *Sunshine* and Icarus II hurtle ever faster to their destination, the plot's twists lean ever more towards the demands of cinema than those of solid science. But you don't go to a blockbuster film to be entertained by two hours of fastidiously researched plausibility: wall-to-wall action, searing special effects, a messianic score and fine acting are generally more than adequate compensation. *Sunshine* has these in abundance.

And yet this film has only one true star: the Sun's yellow–orange plasma, a constant, hypnotic backdrop shown off to full advantage in an early scene as the ship's crew watches the tiny black silhouette of Mercury transit across its shimmering face. At a time when we are ever more concerned with Earthbound threats to humanity's survival, such images are an awe-inspiring tribute to a power of nature that will, in time, be the death of Earth. Five billion years' hence, however, we probably shan't be around to rage against the dying of the light. ■

Richard Webb is *Nature*'s physics News & Views editor.

Staring at the Sun: mankind must try to save our star in the film *Sunshine*.

TWENTIETH CENTURY FOX

Unity from conflict

Could the evolution of multicellular life have been fuelled by conflict among selective forces acting at different levels of organization?

Paul B. Rainey

The transition from unicellular to multicellular life brought with it division of labour, development and a panoply of animal and plant forms. But despite its pivotal role in the history of life, the evolution of multicellularity is poorly understood. Watching the *de novo* emergence of primordial groups in bacterial populations in the lab has led me to think about mechanisms by which higher levels of complexity could emerge.

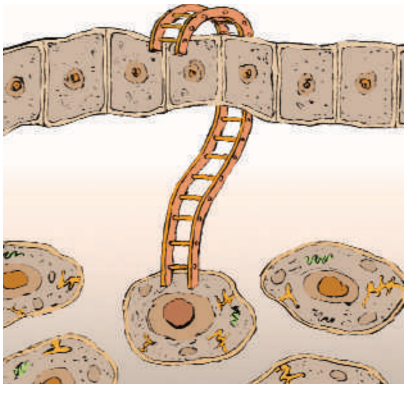
In multicellular organisms, lower-level entities (cells) have relinquished their ability to reproduce as independent units and instead replicate exclusively as part of the larger whole. The challenge for biologists is to explain how a population of cells becomes a single entity capable of self-reproduction; that is, how an evolutionary transition that involves a shift in the level at which selection operates, from individual entities to groups of entities, can occur.

Existing theory predicts a central role for cooperation. This prediction is borne out by experiments using bacterial populations in which the evolution of simple undifferentiated groups has been observed. Groups arise from mutant cells that overproduce a cell–cell glue; the failure of cells to separate at cell division leads to the formation of a surface-colonizing mat. The emergence of these mats — which form at the air–liquid interface, an environment that is replete in oxygen — is a triumph of selection acting at the higher level. Glue production is costly to individual cells, but the trait spreads because the group reaps a reward (oxygen) that is denied to cells within the broth.

Although the evolutionary emergence of groups is itself a profound event, such groups are typically short-lived. The reason is simple: selection continues to act at the level of individual cells. In so doing, selection favours mutant types that cheat; that is, cells that no longer produce the glue, but nonetheless take advantage of the benefit that accrues from being part of the group. In the absence of any mechanism to repress or punish them, cheaters prosper — ultimately weakening the fabric of the mat to the point where it collapses.

Mechanisms that repress cheaters are thus integral to the success of multicellular organisms, but their evolution is a paradox. Cheater control is unlikely to evolve in the absence of selection acting strongly at the group level. The catch-22 is that selection is powerless to act at the group level because newly emerged groups are incapable of differential reproduction. Although individual microbes within the group continue to divide, groups as a whole are not able to produce offspring; they are, like soma (the non-reproductive cells of a multicellular organism), an evolutionary dead-end.

What, then, represents a plausible route to multicellularity? Cooperation is a necessary innovation but it is not the only one. Also required is group reproduction and a way to minimize the disruptive effects of cheats. Moreover, these three major events need to occur more or less simultaneously.



J. KAPUSTA/IMAGES.COM

Although the emergence of these three traits in quick succession seems an impossibly difficult challenge for evolution, alternative scenarios in which multicellularity arises in the absence of higher-level selection simply don't make sense. What follows is one plausible scenario.

Evolutionists are increasingly aware of the importance of a multilevel-selection perspective — a way of making sense of the effects of selection at different levels of biological organization. Integral to this perspective is the recognition that processes occurring at different levels (for example at the level of the cell or of the organism), and at different times, are connected: selection at the higher level has consequences for the lower level and vice versa. Although the consequences can be negative, it is just possible that conflicts between selective forces might fuel, rather than hinder, the transition to higher levels of complexity.

Being unable to replicate itself, a newly emerged group of cells is analogous to the soma. The cheater, however, is loosely analogous to a primordial germ line. Selection therefore favours the evolution of cheats, which although destructive to the group, may act as the seeds for the emergence of new groups. In fact, in our experimental populations the cheats are able to swim away from the group. Moreover, because the number of cheats (germ cells) that a group leaves depends on the rate at which cheats arise, selection acting at the higher 'group' level will optimize the rate at which cheats evolve.

Imagine one group in which the mutation rate to cheat is high and another where the rate is low. A group comprised of few cooperating cells will rapidly be destroyed by cheats; moreover, once the group is destroyed, cheats no longer prosper. So the group that has a high rate of cheating will leave few offspring relative to a group that manages to produce a substantial group-level structure before cheats arise. Selection acting on the rate of cheating thus becomes a mechanism for cheater control. More significantly though, by equating cheats with germ cells, primordial groups are no longer an evolutionary dead-end. Selection can act at the higher level, but the process occurs via effects wrought at the lower level.

Implicit in this model is a capacity for the cheat to repeatedly switch between group-forming and selfish types by compensatory mutation. This may not always be possible, but in our populations of microbes this capacity does exist. Clearly, little progress towards a more complex group can occur without the eventual evolution of a developmentally regulated mechanism for switching between group and cheat (soma and germ line). But this is by no means beyond the capacity of evolution — given an appropriate selective environment. ■

Paul B. Rainey is at the School of Biological Sciences, University of Auckland, Private Bag 92019, Auckland, New Zealand.

FURTHER READING

Maynard Smith, J. & Szathmáry, E. *The Major Transitions in Evolution* (Freeman, Oxford, 1995).
Michod, R. E. *Darwinian Dynamics: Evolutionary Transitions in Fitness and Individuality* (Princeton Univ. Press, 1999).
Okasha, S. *Evolution and the Levels of Selection* (Oxford Univ. Press, 2006).
Rainey, P. B. & Rainey, K. *Nature* **425,** 72–74 (2003).

For other essays in this series, see http://nature.com/nature/focus/arts/connections/index.html

NEWS & VIEWS

NEUROSCIENCE

Controlling neural circuits with light

Michael Häusser and Spencer L. Smith

Two light-sensitive proteins from unicellular organisms have been harnessed to rapidly activate or silence neurons. This optical remote control allows precise, millisecond control of neural circuits.

The prospect of direct brain control has long been a dream of philosophers, science-fiction writers and dictators. Neuroscientists have also indulged in this dream, because observation alone is insufficient to find out how the brain works. Rather, it is necessary to manipulate specific sets of neurons, stimulating or silencing them, to determine their role in brain function. A powerful approach for controlling neural circuits is now described by Zhang and colleagues on page 633 of this issue[1].

The ideal tool for manipulating the activity of neurons would allow targeting of genes in specific neurons, be innocuous in the absence of stimulation, be fully reversible, have a rapid onset and offset, allow bidirectional control, and act at a distance. Such a tool would allow the experimenter to effortlessly manipulate and replay natural patterns of activity in neural circuits.

The traditional approach for controlling neural circuits uses the trusty stimulation electrode, wielded by generations of distinguished neurophysiologists from Charles Sherrington to John Eccles to Bernard Katz. This device generates a local electric field that is highly effective at triggering action potentials in neurons — the wave of electrical depolarization, or 'spike', that ultimately activates the release of neurotransmitter at the synapses between neurons. Although the timing of electrode stimulation is very precise, its specificity and spatial control are poor because all neurons within range are activated. Furthermore, silencing neural activity is as necessary as stimulating it when deciphering brain function, as it is useful to functionally remove cells from a circuit, particularly those that exhibit substantial spontaneous activity. However, silencing is difficult to achieve using a stimulation electrode.

An alternative to electrical stimulation has been the use of chemical activation or inactivation of neurons, by directly applying excitatory and inhibitory neurotransmitters. Recently, 'photo-uncaging' of caged neurotransmitters has made this approach more sophisticated because the use of localized, patterned light yields higher spatial resolution. But the temporal and spatial resolution remains limited by the properties of the cage and the diffusion of neurotransmitter, and specificity cannot be guaranteed. Refinements of this approach, such as two-photon uncaging[2] or restricting receptors for caged neurotransmitters to specific neurons[3], have provided improvements. Furthermore, new genetic tools for silencing neurons[4–6] and specific neural connections[7] have helped to address the specificity problem, but at the expense of temporal resolution.

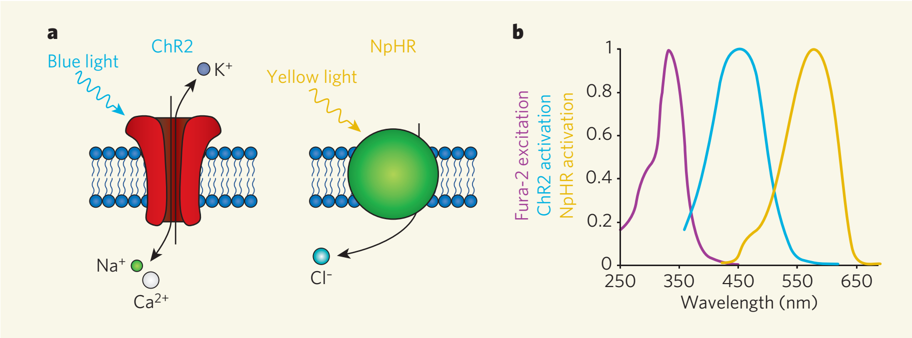


Figure 1 | Light activation of the membrane proteins ChR2 and NpHR. a, Blue light activates ChR2, a cation channel, which depolarizes the cell and causes action potentials. Yellow light activates NpHR, a chloride pump, which causes hyperpolarization and prevents action potentials. **b**, The absorption spectra for these two light-activated proteins are sufficiently separated to ensure that one protein can be activated by light without also activating the other. Furthermore, there is plenty of spectral bandwidth left over to use a fluorescent indicator of intracellular calcium (such as fura-2), which can be used to report neural activity optically.

A pair of light-activated membrane proteins present in nature have now come to the rescue. Zhang *et al.*[1] — a collaboration between neuroengineers at Stanford University and membrane biophysicists at the Max Planck Institute of Biophysics in Frankfurt — have harnessed the power of these proteins to genetically engineer neurons to be controlled by light. In earlier work[8] they used a light-activated ion-channel protein called channelrhodopsin-2 (ChR2) from green algae (Box 1, overleaf). When neurons were genetically engineered to express ChR2, blue light rapidly and selectively triggered action potentials in the neurons (Fig. 1a). This at last gave neuroscientists a genetically encodable, remote-control 'on' switch for neurons.

But that is only half of the story. In the new paper[1] the same team presents the all-important complement to ChR2: *Natronomonas pharaonis* halorhodopsin (NpHR; Box 1). The NpHR protein generates a chloride flux when activated by light. When it is expressed in neurons, activation of this protein by yellow light causes rapid and reversible hyperpolarization, thus preventing spikes (Fig. 1a). As with ChR2, the effect of NpHR is fast and rapidly reversible, despite being an ion pump instead of an ion channel. This rapid-onset hyperpolarization is a dramatic advance over a previous genetically encodable optical membrane probe that relied on the activation of a slower chemical cascade[9].

Thanks to spectacularly good fortune, the excitation spectrum of NpHR is neatly shifted with respect to that of ChR2 (Fig. 1b), such that coexpression of both proteins in the same neuron allows different wavelengths of light to be used to selectively activate or inactivate the same cell. In effect, the system provides the experimenter with a two-knob remote control for increasing or decreasing the activity of specific neurons using different colours of light. A former member of the collaboration has just published an independent study reporting the same technology applied to cultured neurons[10].

To demonstrate the power of the approach, Zhang *et al.*[1] used it in combination with a fluorescent calcium indicator to simultaneously monitor calcium signals associated with activity and manipulate that activity in the same neurons. Activation of ChR2 with

blue light caused large increases in calcium, but when both blue light and yellow light were presented at the same time, no calcium increase was observed. This shows that the ChR2–NpHR system can be combined with fluorescence-based methods for monitoring activity, allowing light to be used to both observe and control neuronal activity. Future studies could employ a fluorescent protein reporter of activity[11], in place of the dye, to create a fully genetically encodable battery of tools for exploring and mapping neural circuits. In a tantalizing taster of the future prospects of the technique, Zhang *et al.* genetically targeted NpHR and ChR2 to specific types of neuron in the worm *Caenorhabditis elegans*. The results showed that it is possible to rapidly and reversibly silence or activate a particular neuronal subtype with light (Fig. 2a), and simultaneously observe the effect on behaviour.

This is a stunning advance. Nevertheless, there is room for improvement. The cation channel of ChR2 allows Ca^{2+} entry into the cell, which triggers many other signalling pathways that could confound some experiments. Although NpHR could be used to mimic inhibition of synapses (for example locally, in the neuronal dendrites or at the spike-initiation site in the axon), it works by activating a pump, not a channel, and thus its biophysical mechanism of action is different. It also remains to be seen whether the ChR2–NpHR system allows efficient two-photon excitation[12], which would allow precise targeting of light deep within brain tissue, or subcellular targeting of ChR2–NpHR activation (for example to axonal boutons or dendritic spines). Even with these issues, which will probably be addressed by protein engineering, the overwhelming advantages of this new approach should revolutionize the field.

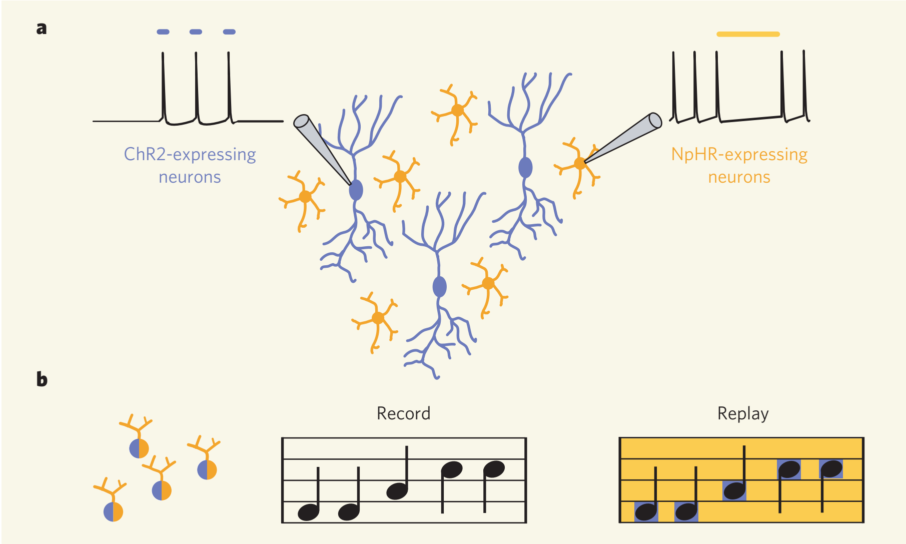


Figure 2 | Remote control of designer neurons[1]. **a,** Selective control of different neuron populations in neural networks. One type of neuron (the normally silent principal cells) is genetically engineered to express ChR2 and fire spikes when activated with blue light. Another type (the spontaneously firing interneurons) express NpHR and are silenced with yellow light. **b,** Precise control of neuronal firing in a neural network coexpressing ChR2 and NpHR, here shown as a musical analogy. First, the pattern of activity in the network is recorded (for example using calcium dyes). Next, the identical pattern of activity is replayed into the same populations of cells using yellow light to suppress spurious spikes and blue light to activate spikes. Such an experiment can help to demonstrate causality of patterns of activity for behaviour.

Proving causality is often an elusive goal in neuroscience, particularly when drawing connections between neural activity and behaviour. The ChR2–NpHR system offers an elegant solution for proving necessity and sufficiency, which together allow causality to be inferred. Patterns of neural activity during stimulus-evoked behaviour could be recorded using existing optical methods. Then, to demonstrate sufficiency, these same patterns could be played back to the neuronal circuit, using ChR2 and NpHR, to elicit the same behaviour in the absence of the original stimulus (Fig. 2b). Changing the precise temporal pattern of spikes in such a replay experiment could also allow one to assess the importance of precise spike timing in behaviour. Subsequently, NpHR activation alone could silence stimulus-evoked neuronal activity to block behaviour, thus demonstrating necessity.

Beyond the laboratory, the ChR2–NpHR system has potential for treating neurological and psychiatric disorders in humans. For example, diseases such as epilepsy, which involves hyperactivity or altered excitation–inhibition balance in neurons, might be ameliorated by activating NpHR and/or ChR2 targeted to specific cell types. With implanted fibre-optic stimulation, these optical tools could also provide a more specific and durable form of deep-brain stimulation, which has shown some success in the treatment of Parkinson's disease. Finally, one can envisage using this approach to create new types of optical neural prosthetic, such as in the retina[13] or the sensory or motor cortex, for improving sensory perception and control of movement. Our new mastery of neural circuits using light might thus eventually allow us to master our own brains — and behaviour. ■

Michael Häusser and Spencer L. Smith are at the Wolfson Institute for Biomedical Research and Department of Physiology, University College London, Gower Street, London WC1E 6BT, UK.
e-mail: m.hausser@ucl.ac.uk

Box 1 | A tale of two opsins

ChR2 and NpHR are both opsins that are related to rhodopsin, the light-sensitive protein in mammalian vision. But unlike rhodopsin, which activates a chemical cascade that eventually results in neural activity, ChR2 and NpHR are self-contained, light-sensitive proteins that produce a direct electrical signal (depolarization or hyperpolarization, respectively).

Like all opsins, though, they require a vitamin-A-based chromophore cofactor in order to be activated by light. Conveniently, mammalian neural tissue normally contains enough of this cofactor, which means the experimenters did not need to add any.

ChR2 was discovered in the green alga *Chlamydomonas reinhardtii*, where it triggers photo-orientation movements that allow the alga to find optimal conditions for photosynthesis. Its potential as a light-sensitive depolarizer for cells did not escape the attention of the Max Planck group that sequenced the gene[14] (and which led to their collaboration on Zhang and colleagues' study[1]).

The most commonly studied variant of *C. reinhardtii* was isolated in Massachusetts. NpHR, on the other hand, was found thousands of miles away in a dry salt lake in the Sahara desert. In this extreme environment, the archaeum *Natronomonas pharaonis* fills itself with chloride using a pump (NpHR), probably for generating an electrochemical gradient for proton entry that in turn is used to drive ATP synthesis for energy production. This amazing protein can pump chloride into the cell to concentrations 50 times higher than that typically found in neurons.

Apart from some minor optimization of nucleic-acid sequence to increase their levels of expression in mammalian neurons, the ChR2 and NpHR proteins were essentially used in their natural forms for the present work[1]. They will be attractive targets for protein engineers eager to improve and expand on these proteins' already useful properties. **M.H. & S.L.S.**

1. Zhang, F. *et al. Nature* **446,** 633–639 (2007).
2. Matsuzaki, M. *et al. Nature Neurosci.* **4,** 1086–1092 (2001).
3. Lima, S. Q. & Miesenböck, G. *Cell* **121,** 141–152 (2005).
4. Slimko, E. M., McKinney, S., Anderson, D. J., Davidson, N. & Lester, H. A. *J. Neurosci.* **22,** 7373–7379 (2002).
5. Lechner, H. A., Lein, E. S. & Callaway, E. M. *J. Neurosci.* **22,** 5287–5290 (2002).

6. Burrone, J., O'Byrne, M. & Murthy, V. N. *Nature* **420,** 414–418 (2002).
7. Karpova, A. Y. *et al. Neuron* **48,** 727–737 (2005).
8. Boyden, E. S. *et al. Nature Neurosci.* **8,** 1263–1268 (2005).
9. Li, X. *et al. Proc. Natl Acad. Sci. USA* **102,** 17816–17821 (2005).
10. Han, X. & Boyden, E. S. *PLoS ONE* **2,** e299 (2007).
11. Miyawaki, A. *Neuron* **48,** 189–199 (2005).
12. Denk, W., Strickler, J. H. & Webb, W. W. *Science* **248,** 73–76 (1990).
13. Bi, A. *et al. Neuron* **50,** 23–33 (2006).
14. Nagel, G. *et al. Proc. Natl Acad. Sci. USA* **100,** 13940–13945 (2003).

ATTOPHYSICS

Tunnel vision

Jonathan P. Marangos

The tunnelling of a bound electron out of an atom in a laser field is a well-known quantum-mechanical process. But it happens very quickly, and it takes some fast work with X-rays and lasers to see it in action.

Advances towards making measurements with attosecond time resolution have caused much excitement in recent years. An attosecond is 10^{-18}s, and this is the timescale over which electron states in matter evolve. Such evolution drives many processes that are of interest to quantum scientists, among them chemical reactions, changes in material structures, optical responses and quantum coherence. Elsewhere in this issue, Ferenc Krausz and colleagues (Uiberacker *et al.*)[1] extend a technique that brings us within sight of these fast-moving phenomena.

Measurements in the attosecond domain are based on the use of high-intensity laser pulses that comprise just a few wave-cycles of electrical field and have a precisely defined form[2]. This strong, oscillating electrical field can be used to extract an electron from an atom, in a quantum effect termed tunnel ionization, and then drive it back in with increased energy. This process is signalled by a short burst of coherent X-rays, typically a few hundred attoseconds long[3], that is perfectly synchronized with the laser field's cycle. Such X-ray bursts have been used in various ways to investigate, for example, the movements of protons in small molecules[4], and the decay of states in atoms where several electrons have been excited[5].

Krausz and his colleagues at the Max Planck Institute for Quantum Optics in Garching, Germany, were pioneers of these methods. They use a precisely controlled laser field that has a cycle time of around 2,700 attoseconds to produce, in a gas jet, a synchronous X-ray burst some 250 attoseconds long and with a photon energy of around 90 electronvolts. This X-ray burst is delivered to the sample to be measured, followed, after a carefully controlled delay, by the original laser pulse. In previous experiments[5], the researchers had clocked the emission time of X-ray-excited electrons with a precision of around 100 attoseconds by measuring the effect of the delayed laser field on the energy distribution. This set-up, termed attosecond streaking spectroscopy, worked well for excitation processes in which an electron is eventually ejected. But in many cases, the electron does not leave the sample. A new technique was needed to measure the dynamics of these more deeply bound states.

In their latest paper (page 627)[1], the researchers describe how they have used the same basic tools to look into such deep states in neon (Ne) atoms (Fig. 1). The 90-eV photons of the X-ray burst had enough energy to free one electron from the atom immediately and excite a second to a higher, but still bound, energy level, creating an excited ionic state, Ne^{+*} (Fig. 1a). When the subsequent laser electrical field is near a maximum, it can strongly distort the potential that binds the electron. The excited electron encounters a finite barrier through which, with high probability, it can tunnel to freedom[6]. An appropriate maximum of the laser field occurs twice per optical cycle (Fig. 1b). The signal of Ne^{2+} ions thus accumulates over time in steps.

The accumulated yield of Ne^{2+} ions depends on the delay between the X-ray burst and the laser pulse. The number of field maxima that can contribute to the Ne^{2+} signal, and so the Ne^{2+} yield, changes in steps as the delay is varied. The steepness of the steps indicates the duration of the tunnel ionization. This behaviour shows that the X-ray excitation and ionization processes of these excited states are indeed very rapid, and must occur in a combined time of less than 300 attoseconds. Although this temporal confinement of the laser-induced ionization had long been suspected, it had not been directly confirmed before. The measurements also show that tunnel ionization becomes the principal ionization process at surprisingly low intensities.

Because the tunnelling from the bound states is so rapid and so exactly synchronized to the X-ray burst, the yield of ions from this process is a suitable probe of their electron dynamics.

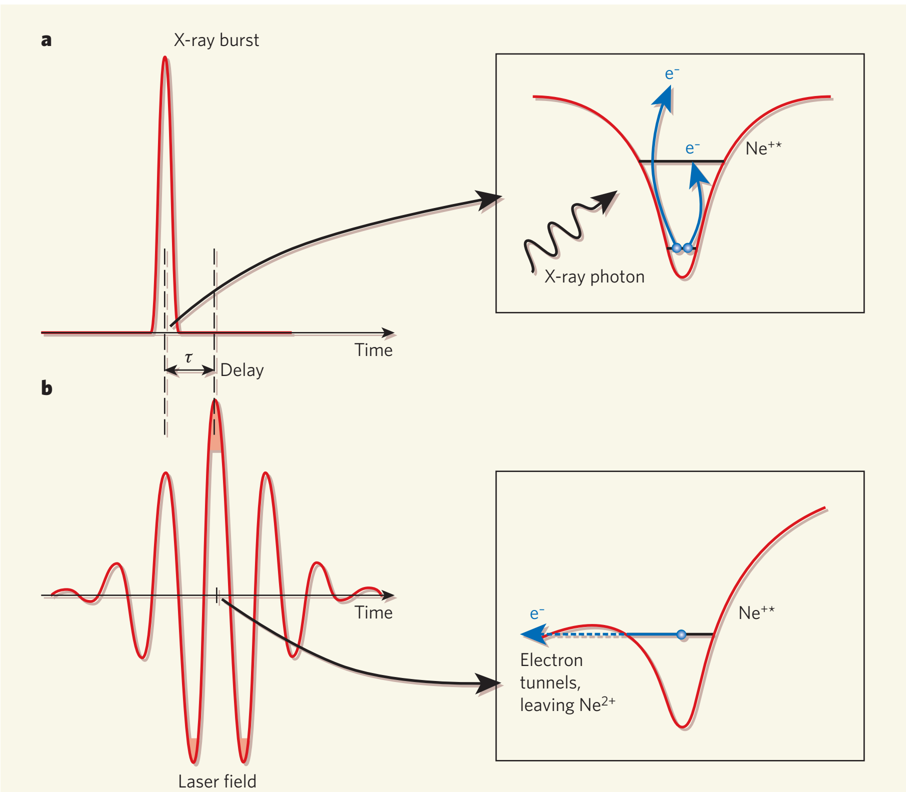


Figure 1 | Attosecond probe. In the experiments of Uiberacker *et al.*[1], the X-ray burst and the laser electrical field have a relative delay, τ, between their centres that can be varied. **a**, The X-ray burst excites two electrons in a neon (Ne) atom. One of these leaves directly, and the other is excited into a bound state of the Ne^{+} ion, Ne^{+*}. **b**, The following laser field can, when near one of its maxima (in one of the shaded regions), significantly perturb the Ne^{+} ion potential, allowing the second electron to tunnel away quantum-mechanically, resulting in a significant degree of ionization to Ne^{2+}. As the delay τ is changed, the X-ray burst scans across the laser-field profile. The number of field maxima that can contribute to the Ne^{2+} signal, and so the Ne^{2+} yield, changes in steps as τ is varied.

The authors argue that further measurements with improved precision will allow tunnelling to be characterized more fully. Such measurements will require X-ray pulses significantly shorter than those of a few hundred attoseconds that are presently available. Once this hurdle has been surmounted, the tunnelling method could be used to probe how electrons behave in a wide range of atomic and molecular systems with attosecond precision.

In the meantime, Uiberacker *et al.* turn their attention[1] to the significantly slower processes associated with the decay of excited states in the xenon ions Xe^{2+} and Xe^{3+}. Using their laser-field tunnelling method to produce Xe^{4+} ions, they have been able to measure cascaded population transfer processes among the electron states of ions that have lifetimes of as little as 6 femtoseconds (6×10^{-15} s). They thus already show the potential of their method to extend and increase the precision of findings established in earlier spectroscopic studies. ■

Jonathan P. Marangos is in the Blackett Laboratory, Department of Physics, Imperial College London, London SW7 2BW, UK.
e-mail: j.marangos@imperial.ac.uk

1. Uiberacker, M. *et al. Nature* **446,** 627–632 (2007).
2. Baltuska, A. *et al. Nature* **421,** 611–615 (2003).
3. Sansone, G. *et al. Science* **314,** 443–446 (2006).
4. Baker, S. *et al. Science* **312,** 424–427 (2006).
5. Drescher, M. *et al. Nature* **419,** 803–807 (2002).
6. Oppenheimer, J. R. *Phys. Rev.* **31,** 66–81 (1928).

PLANT BIOLOGY

Sticking with auxin

Tom Guilfoyle

Auxin is one of the main agents that regulate plant growth and development. Intricate crystallographic studies reveal how this hormone acts as a 'molecular glue' in mediating substrate–receptor interactions.

Research into plant hormones has been coming on apace in the past few years. The receptors for these hormones that have been identified, including that for auxin, have turned out to be surprisingly different from the receptors for animal hormones[1]. On page 640 of this issue, Tan *et al.*[2] take the story further — their first-of-a-kind crystallographic studies provide more revelations about auxin perception and the auxin receptor.

In 2005, a receptor for auxin (or indole-3-acetic acid, IAA) was identified as the F-box protein TIR1, short for 'transport inhibitor response 1'; TIR1 is a component of a cellular protein complex known as SCF^{TIR1} (refs 3, 4). The substrates for TIR1, Aux/IAA repressors, are recruited to the receptor in an auxin-dependent manner and, after binding to TIR1, are degraded. Identification of the TIR1 receptor suggested that auxin perception and the signalling pathway to auxin-regulated gene expression was direct and simple, but it left various questions. Tan *et al.*[2] now describe crystal structures of a TIR1 complex that reveal how auxins fit into a surface pocket of TIR1 and enhance the binding of Aux/IAA repressors to TIR1.

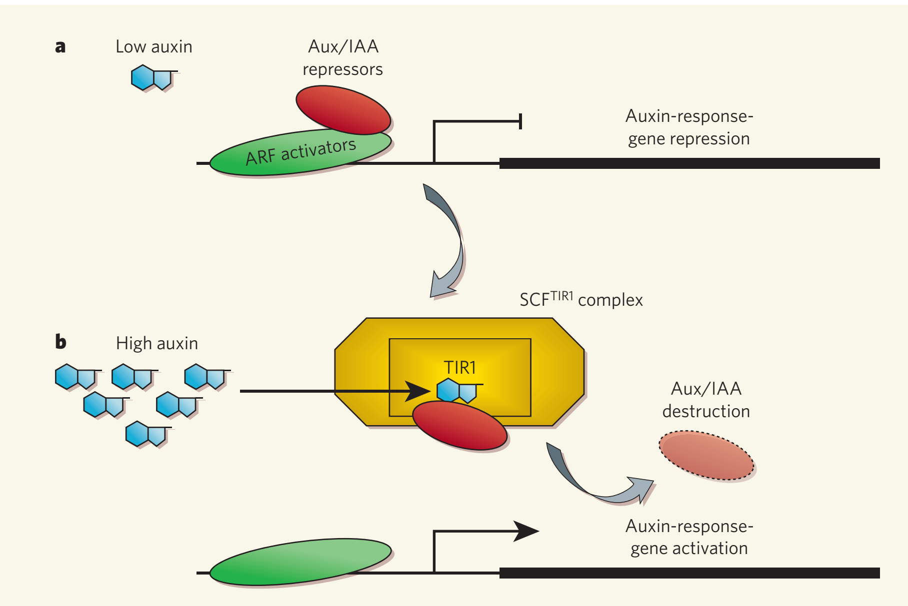


Figure 1 | The auxin signalling pathway. ARF activators bind to auxin-response elements in promoters of auxin-response genes. **a**, When auxin concentrations are low, Aux/IAA repressors associate with the ARF activators (via domains III and IV, see Fig. 2) and repress expression of the genes. **b**, When auxin concentrations increase, auxin binds to the TIR1 receptor in the SCF^{TIR1} complex, leading to recruitment of the Aux/IAA repressors to TIR1. Once recruited to the SCF^{TIR1} complex, the repressors enter a pathway that leads to their destruction and the subsequent activation of the auxin-response genes.

Auxin-regulated gene expression triggers most of the processes controlled by this plant hormone. Many auxin-induced genes are regulated by the interplay of two classes of gene-transcription factors, auxin-response factors (ARFs) and the Aux/IAA repressors[5] (Fig. 1). ARFs bind to auxin-response promoter elements in auxin-response genes. When auxin concentrations are below a threshold level, Aux/IAA repressors associate with ARF activators and repress the expression of these genes. Conversely, higher levels of auxin lead to destruction of the Aux/IAA repressors, and to activation of the genes[6,7].

Aux/IAA repressors contain four conserved domains, and one of these, domain II, is responsible for the instability of these proteins (Fig. 2a, overleaf). Domain II has a hallmark GWPPV amino-acid motif that is recognized by TIR1 in the SCF^{TIR1} complex[8,9]. IAA binds to TIR1 to enhance the recruitment of the motif to TIR1, as do two synthetic auxins — 1-naphthalene acetic acid (1-NAA) and 2,4-dichlorophenoxy-acetic acid (2,4-D, which is used as a herbicide) — but with different affinities (Fig. 2b).

Tan *et al.*[2] expressed TIR1 complexed with ASK1 (a SCF^{TIR1} adaptor) in a baculovirus–insect system. After demonstrating that auxin enhanced the binding of an Aux/IAA protein to the complex, they obtained crystal structures for the complex alone and for complexes bound to IAA and the two synthetic auxins along with an Aux/IAA peptide containing the GWPPV motif (Fig. 2). The crystal structures showed that the TIR1–ASK1 complex had a mushroom shape, with the leucine-rich-repeat domain of TIR1 forming the cap, and the F-box of TIR1 along with ASK1 forming the stem. A pocket on the top of the TIR1 leucine-rich-repeat domain functions in both auxin binding and substrate recruitment.

It turns out that auxin binds to the bottom of the pocket in a 'promiscuous' binding site that tolerates moderately different planar ring structures (that is, natural and synthetic auxins). The Aux/IAA peptide binds in close proximity to the auxin-binding site in the upper part of the pocket. The GWPPV motif is packed directly against auxin and covers the auxin-binding site. This is thought to trap auxin in the binding pocket until the Aux/IAA peptide is released and moved along the degradation pathway. The crystal structure also revealed an unexpected moiety in TIR1, which turned out to be a tightly bound inositol hexakisphosphate ($InsP_6$) molecule. $InsP_6$ functions in many

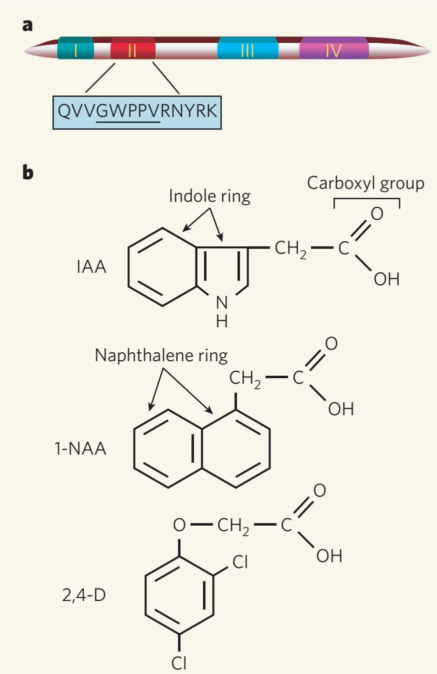


Figure 2 | The main molecular players in the work of Tan *et al.*[2]. a, Depiction of the four conserved domains of Aux/IAA repressors. The synthetic peptide from domain II is sufficient for targeting Aux/IAA to the TIR1 auxin receptor, the core sequence being GWPPV (G, glycine; W, tryptophan; P, proline; V, valine). **b**, The three auxin ligands that the authors crystallized in association with the TIR1 auxin receptor. IAA itself, and the synthetic auxins 1-NAA and 2,4-D, bind to a 'promiscuous' cavity in the receptor with different affinities, but all of them stabilize the interaction between the Aux/IAA repressor and the receptor.

cellular processes, including acting as a phosphate reservoir in plants. Its association with TIR1 appears to be essential for auxin binding and function of the receptor.

The authors compared TIR1 structures that had IAA or the two synthetic auxins in the ligand binding site. IAA binds to TIR1 with the greatest affinity of the three auxins, and binding involves its side-chain carboxyl group as well as its indole ring (Fig. 2b). The synthetic auxins bind to TIR1 in a manner similar to IAA, but with affinities determined by how effectively their ring structures fit into and interact with the promiscuous cavity of the receptor. The auxin herbicide 2,4-D has the weakest affinity for TIR1, having the smallest ring and the least amount of surface contact with the cavity of the receptor. Although 1-NAA has the largest ring and the most surface contact with the cavity, it has only an intermediate affinity for TIR1. The smaller indole ring of IAA does not fit the cavity as well as the naphthalene ring of 1-NAA, but it makes an additional contact with the receptor, which increases its affinity.

Interestingly, the binding of auxin does not induce significant conformational changes in TIR1. Instead, auxin enhances the binding of Aux/IAA substrate to TIR1 by occupying a cavity between substrate and receptor, thus forming a continuous hydrophobic core among ligand, substrate and receptor. The authors characterize this as a 'molecular glue' that effectively strengthens the binding of the substrate to the receptor.

Tan *et al.*[2] have provided the first detailed structure of a plant hormone receptor, but further questions are the inevitable upshot. Do plant hormones such as jasmonic acid bind to and function as molecular glue in TIR1-related F-box proteins[10]? Are there TIR1-like proteins in organisms besides plants that function as hormone receptors or small-molecule sensors? And how interconnected is InsP6 metabolism with auxin signalling and TIR1 activity? Several labs will soon be hard at work in taking these next steps. ■

Tom Guilfoyle is in the Department of Biochemistry, University of Missouri, 117 Schweitzer Hall, Columbia, Missouri 65211, USA.
e-mail: guilfoylet@missouri.edu

1. Bishopp, A., Mahonen, A. P. & Helariutta, Y. *Development* **133,** 1857–1869 (2006).
2. Tan, X. *et al. Nature* **446,** 640–645 (2007).
3. Dharmasiri, N., Dharmasiri, S. & Estelle, M. *Nature* **435,** 441–445 (2005).
4. Kepinski, S. & Leyser, O. *Nature* **435,** 446–451 (2005).
5. Hagen, G. & Guilfoyle, T. J. *Plant Mol. Biol.* **49,** 373–485 (2002).
6. Zenzer, N. *et al. Proc. Natl Acad. Sci. USA* **98,** 11795–11800 (2001).
7. Tiwari, S. B. *et al. Plant Cell* **13,** 2809–2822 (2001).
8. Gray, W. M. *et al. Genes Dev.* **13,** 1678–1691 (1999).
9. Kepinski, S. & Leyser, O. *Proc. Natl Acad. Sci. USA* **101,** 12381–12386 (2004).
10. Xie, D. X. *et al. Science* **280,** 1091–1094 (1998).

EVOLUTIONARY BIOLOGY

Born-again hagfishes

Philippe Janvier

The strange, slimy creatures called hagfishes are of abiding interest to students of vertebrate evolution: just where do they fit in? Investigations of hagfish development take the story forward.

Hagfishes are almost blind, cartilaginous, eel-shaped, marine vertebrates and, most notably, they lack jaws. Their relationships to other major living vertebrate groups — the similarly jawless lampreys, and the jawed vertebrates — remain contentious, and one avenue of investigation is to look to embryonic development for further information. Alas, hagfish embryos that are suitable for such studies have been desperately rare. On page 672 of this issue[1], however, Ota and colleagues describe the first early embryos from hagfishes to have been found since 1930, and report their studies on gene-expression patterns in them*.

Hagfishes are occasional scavengers, and are frequently found inside dead fish. So Linnaeus[2] classified them among the 'intestinal worms', although he noticed that they share some characteristics with lampreys, and thus with vertebrates. They then became classified with lampreys as cyclostomes ('rounded mouth') because both possess a jawless mouth armed with retractable horny teeth, and gills enclosed in pouches. Living vertebrates thus fell into two major groups, the cyclostomes and the jawed vertebrates (or gnathostomes), which remained the received view for more than 170 years.

However, biologists progressively noticed that hagfish anatomy and physiology were in many respects more 'simple' or 'primitive' than those of lampreys and jawed vertebrates. Unlike these two groups, hagfishes lack vertebrae, heart innervation, eye lenses, lymphoid tissues, a perfected adaptive immune response, and many other classical vertebrate characters. In these respects, then, they resemble non-vertebrate chordates such as amphioxus or sea-squirts. That meant that there are two possibilities (Fig. 1). First, that hagfishes are 'degenerate' cyclostomes — that is, they have lost several vertebrate characters that lampreys and jawed vertebrates retain (unless independently acquired, which is unlikely). Second, that they are actually the most primitive vertebrates, and so are the 'sister group' of all other vertebrates[3]. Phylogenies based on DNA and RNA sequencing generally support the first hypothesis, but remain ambiguous[4,5].

Palaeontology sometimes settles such conflicts. But it is powerless in this case, because the earliest (300-million-year-old) hagfishes, preserved as soft-tissue imprints, are very similar to living ones. Moreover, the distribution of lampreys and gnathostomes through time suggests that hagfishes had a 'ghost range' of between 50 million and 170 million years during which they were apparently present, but are unrecorded[6]. In such cases, embryonic development may provide clues, on the assumption that early embryos may mirror ancestral conditions. But hagfish eggs found on the sea floor generally contain no visible embryo, probably because their development is very slow and early embryos may be missed[1]. Out of the 150 fertilized eggs found between 1896 and 1930, only a few embryos could be studied, but the techniques then in use did not allow detailed description.

However, among the embryonic hagfish

*This article and the paper concerned[1] were published online on 18 March 2007.

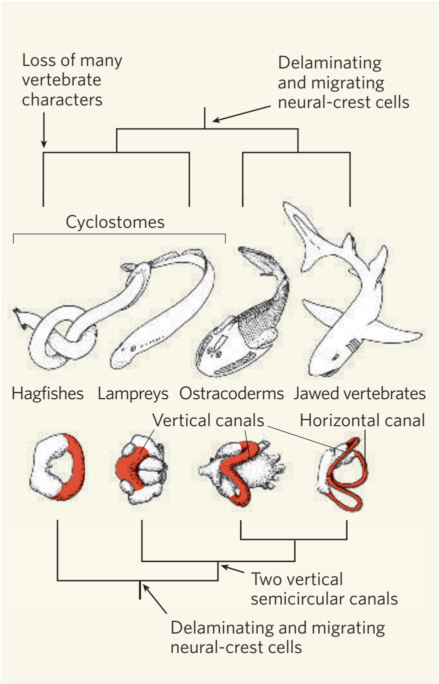


Figure 1 | The hagfish puzzle. Ota *et al.*[1] show that hagfishes have delaminating and migrating neural-crest cells, a feature of vertebrate head development. But it is still not clear whether hagfishes are 'degenerate' cyclostomes (top tree), or are 'primitive' and are the sister group of all other vertebrates (bottom tree). This second hypothesis is supported by many characters shared only by lampreys, jawed vertebrates and jawed-vertebrate precursors, the 480–365-million-year-old jawless ostracoderms. Notably, there are differences in the canals (red) of the inner ear (labyrinth): the hagfish labyrinth possesses a single semicircular canal for balance, whereas all other vertebrates have at least two, vertical, canals.

features described 65 years ago[7] was the odd development of the 'neural crest', a unique vertebrate structure that has possible precursors in other chordates[8,9]. During the embryonic development of lampreys and gnathostomes, the brain and spinal cord form from a dorsal infolding of the ectoderm (the future epidermis of the skin), which then closes into a 'neural tube'. Where the two lips (or crests) of this infolding meet to close the neural tube, some cells of the ectoderm delaminate, becoming free, and then migrate to contribute to essential organs of the vertebrate body, such as the branchial apparatus, part of the braincase, and the dermal bones, scales, teeth, nerve ganglia and pigment cells. The neural crest thus generates a 'new head', which possibly made vertebrates more competitive in the early stage of their evolution[10]. Early studies of hagfish embryos suggested that their neural crests remained in the form of a simple epithelial 'pocket' that did not delaminate and migrate in the same way as in other vertebrates.

Ota *et al.*[1] have been both lucky and ingenious in finding the right style of aquarium where the shy and capricious hagfishes can thrive and spawn. But they were also helped by local fishermen, who knew of the secret and seasonal behaviour of their subjects. Thanks to these efforts, Ota *et al.* can now show that the epithelial pocket is a mere artefact, and that the neural-crest cells of hagfishes delaminate and migrate like those of all other living vertebrates. With this discovery, hagfishes become more 'conventional' vertebrates. And the increased understanding of hagfish spawning habits will make access to their embryos much easier, and thereby help in understanding why certain features of their anatomy and physiology are so odd and apparently primitive.

For example, the labyrinth (inner ear) of vertebrates comprises bow-shaped semicircular canals that ensure balance (Fig. 1). Hagfishes have only one such canal, with sensory ampullae forming a swelling at each end, hence the doughnut shape of their labyrinth that is already visible in early embryos[1] and has long been regarded as primitive. Lampreys and gnathostomes have two vertical canals, each with one ampulla, and gnathostomes have an additional horizontal canal. There is no intermediate between these three conditions, even among fossil 'stem' gnathostomes, the jawless ostracoderms, whose labyrinth is lamprey-like.

Which of these conditions is primitive for vertebrates? Is the hagfish condition derived from that of lampreys by fusion of the two vertical canals? Two genes, *Otx1* and *BMP2*, have proved to be involved in the development of semicircular canals[11,12]. But the morphogenetic control of such remarkably conserved anatomical features also involves complex pathways that regulate their development. Further analyses of the developmental genetics of hagfish embryos might enable us to discover whether hagfish anatomy is primitive or degenerate, and may help in reconstructing the theoretical common ancestor to all vertebrates. ■

Philippe Janvier is at the Muséum National d'Histoire Naturelle, UMR 5143, CNRS, 8 Rue Buffon, Paris 75005, France.
e-mail: janvier@mnhn.fr

1. Ota, K. G., Kuraku, S. & Kuratani, S. *Nature* **446,** 672–675 (2007).
2. Linnaeus, C. *Systema Naturae. Regnum Animale* 10th edn (1758).
3. Løvtrup, S. *The Phylogeny of Vertebrata* (Wiley, New York, 1977).
4. Furlong, R. F. & Holland, P. *Zool. Sci.* **19,** 593–599 (2002).
5. Gursoy, H. C., Koper, D. & Benecke, B. J. *J. Mol. Evol.* **50,** 456–464 (2000).
6. Gess, R. W., Coates, M. I. & Rubidge, B. S. *Nature* **443,** 981–984 (2006).
7. Conel, J. L. *J. Comp. Neurol.* **76,** 191–215 (1942).
8. Le Douarin, N. M. & Kalcheim, C. *The Neural Crest* (Cambridge Univ. Press, 1999).
9. Jeffery, W. R., Strickler, A. G. & Yamamoto, Y. *Nature* **431,** 696–699 (2004).
10. Gans, C. & Northcutt, R. G. *Science* **220,** 268–274 (1983).
11. Mazan, S., Jaillard, D., Baratte, B. & Janvier, P. *Evol. Dev.* **2,** 186–193 (2000).
12. Chang, W., Brigande, J. V., Fekete, D. M. & Wu, D. K. *Development* **131,** 4201–4211 (2004).

PALAEOMAGNETISM

A more ancient shield

David J. Dunlop

Earth's magnetic field has protected our atmosphere from erosion by the solar wind ever since it started up. Silicate crystals from some of Earth's oldest rocks date that event to more than 3 billion years ago.

Most planetary magnetic fields originate from the movement of electrically conducting fluid in the planet's interior: Earth's own magnetic dynamo lies deep in its hot, convecting core of liquid iron. The resulting global magnetic field protects Earth's atmosphere from depletion by the solar wind, a continuous, but highly variable, stream of charged particles ejected from the Sun. On page 657 of this issue, Tarduno *et al.*[1] push back the start-up date of this protective magnetic shield, helping to constrain ideas about when Earth's mantle and outer and inner cores differentiated.

Just how different a planetary body looks without a global magnetic field is shown by our two nearest neighbours, the Moon and Mars. Both had magnetic fields 4 billion years ago. Between 4 billion and 3 billion years ago, however, their magnetic fields died[2], probably because their cooling planetary cores no longer convected vigorously enough. Without shielding from the solar wind, they lost most of their atmospheres. Giant impact craters on Mars that are 3.9 billion years old are unmagnetized, testifying to an early shut-off of the martian field. Samples brought back from the Moon record an initially Earth-like field strength that declined between 3.9 billion and 3.1 billion years ago[2].

Tarduno and colleagues show[1] that Earth's magnetic field was in existence 3.2 billion years ago, 500 million years before the previous earliest proposed start-up date[3], and at a time when the Moon's field was still present, but very much in decline. Using rocks to date ancient magnetic fields is, however, a tricky business. The magnetization of most Precambrian rocks — those older than about 540 million years — has changed as a result of geological processes that involve heating to 250 °C or more. Magnetic material can also be altered chemically, or new generations of material

added when iron leaches from other minerals.

Tarduno *et al.* minimize such chemical and thermal remagnetization effects by analysing the silicate minerals quartz and microcline feldspar in rocks from the huge dome of ancient rock known as the Kaapvaal craton in South Africa (Fig. 1). These minerals are not themselves magnetic, but they contain traces of magnetite, an iron oxide. The silicate jacket of the magnetite crystals protects them from chemical alterations, and their tiny size (a few hundred nanometres across) places them in a 'single-domain' state, in which neighbouring atomic spins (spin determines the direction of a magnetic field) are locked parallel. It is much more difficult for the magnetic moment of these single-domain crystals to change than it is for that of iron-rich silicates, such as hornblende, which precipitate larger magnetite crystals containing multiple, changeable magnetic domains[4].

The authors use a laser to heat millimetre-sized silicate grains in a stepwise fashion. This technique is well established in $^{40}Ar/^{39}Ar$ geochronology[5], but not in palaeomagnetism. It allows fast, uniform heating that erases any heating-induced magnetization changes since the mineral's formation. Between about 400 °C — when, according to theory[2,6], the effects of subsequent remagnetizations are eliminated in the laboratory — and magnetite's Curie point (the temperature above which magnetization is removed altogether) at 580 °C, the single-domain magnetite inclusions will represent the original magnetic field in which the rocks cooled 3.2 billion years ago.

In Tarduno and colleagues' analysis[1], the direction of ancient magnetization, as well as its intensity, are defined with unusually low variability over the crucial 400–580 °C interval: all 10 samples from one rock formation and 11 of 15 samples from the other gave trustworthy ancient field strengths. The identification of this magnetization with the mineral's original magnetization is also strong: the quartz and microcline magnetizations in the 400–580 °C range have directions that are distinct from those of known subsequent remagnetizations. Disagreement between samples mainly stems from 180° reversals. These flips of the magnetic-field axis, with an interchange of north and south poles, have been a regular feature of Earth's recent magnetic past, occurring about once every million years. It seems that this pattern was similar in Earth's early history.

The authors also calculate an ancient field strength that is very similar to the field's strength today. But they tentatively suggest that their raw estimates might be as much as two times too high. A large body of rock requires 10 million years or more to cool from 580 °C to 400°C; small laboratory samples cover this temperature range in minutes. The longer timescale promotes more efficient recording of the magnetic field, making the rock's natural magnetization higher than it is found to be in laboratory tests. Direct evidence of what cooling-rate correction to apply is lacking, however; all experiments are necessarily short in geological terms[7,8]. Halving the raw field strengths might well be an overcorrection. It should be no wonder that Earth's early global magnetic field might have been as strong as today's: the early Earth's core was hotter than it is now, and would have convected just as vigorously.

Tarduno and colleagues' results imply that the differentiation of Earth's solid inner core began no later than 3.2 billion years ago. Core convection driven by temperature differences has difficulty generating a magnetic dynamo on its own; compositional differences between the inner and outer cores are also needed[9]. Thermodynamic model calculations[9] give a most likely starting date of 3.5 billion years ago for inner-core solidification, which is entirely compatible with Tarduno and colleagues' measurements.

Is it possible to date Earth's magnetic field back further? Rocks almost 4 billion years old occur in South Africa, Australia and Canada. Most, however, have been heavily reworked and are unlikely to have preserved pristine information on Earth's early magnetic field. One exception, ancient lava found in the Komati formation of South Africa, was believed to preserve a 3.5-billion-year-old field record[10], but its magnetization is now argued to be of chemical, not thermal, origin[11]. This makes its age uncertain, and any estimate of field intensity is compromised: chemical magnetization is a relatively inefficient recorder of field strength.

Gases removed from Earth's atmosphere by the early solar wind and implanted in lunar mineral grains are an indirect indication of when Earth's field might have started up[12]. Nitrogen and argon isotopes present in the grains could have been stripped from Earth's atmosphere between 3.9 billion and 3.8 billion years ago, at a time when there was no shielding global field[12]. If this inference is correct, Earth's field started up no earlier than 3.9 billion years ago. This line of reasoning has yet to be tested by isotopic analyses of farside lunar soils, which have not been exposed to any flux of ions from Earth.

The onset of a magnetic shield around Earth is a burning question, not least because bombardment by energetic extraterrestrial particles, such as those of the solar wind, might have caused genetic mutations that influenced the evolution of early life on Earth. The best evidence would be a direct record of the early Earth's field, or its absence, in rocks 3.5 billion years old or older. Such evidence will not be easy to come by, but work such as that of Tarduno *et al.*[1] is another step in the right direction. ■

David J. Dunlop is in the Department of Physics, University of Toronto, Toronto, Ontario M5S 1A7, Canada.
e-mail: dunlop@physics.utoronto.ca

J. A. TARDUNO

Figure 1 | Rock of the dome. **The Dalmein pluton was one of two igneous intrusions in the Kaapvaal craton, a dome of rock covering an area of more than 1 million km^2 in South Africa, whose magnetism was investigated by Tarduno *et al.*[1]. The Kaapvaal craton represents some of the oldest exposed rock on Earth, with ages of between 2.5 billion and 3.6 billion years.**

1. Tarduno, J. A., Cottrell, R. D., Watkeys, M. K. & Bauch, D. *Nature* **446,** 657–660 (2007).
2. Dunlop, D. J. & Özdemir, Ö. *Rock Magnetism: Fundamentals and Frontiers* (Cambridge Univ. Press, 1997).
3. Dunlop, D. J. & Yu, Y. *AGU Geophys. Monogr. Ser.* **145,** 85–100 (2004).
4. Dunlop, D. J., Zhang, B. & Özdemir, Ö. *J. Geophys. Res.* **110,** B01103 doi:10.1029/2004JB003095 (2005).
5. Layer, P. W., Hall, C. M. & York, D. *Geophys. Res. Lett.* **14,** 756–760 (1987).
6. Dunlop, D. J. & Buchan, K. L. *Phys. Earth Planet. Inter.* **13,** 325–331 (1977).
7. Bowles, J., Gee, J. S., Kent, D. V., Bergmanis, E. & Sinton, J. *Geochem. Geophys. Geosyst.* **6,** Q07002 doi:10.1029/2004GC000900 (2005).
8. Leonhardt, R., Matzka, J., Nichols, A. R. L. & Dingwell, D. B. *Earth Planet. Sci. Lett.* **243,** 282–292 (2006).
9. Gubbins, D., Alfè, D., Masters, G., Price, G. D. & Gillan, M. *Geophys. J. Int.* **157,** 1407–1414 (2004).
10. Hale, C. J. *Earth Planet. Sci. Lett.* **86,** 354–364 (1987).
11. Yoshihara, A. & Hamano, Y. *Precambr. Res.* **131,** 111–142 (2004).
12. Ozima, M. *et al. Nature* **436,** 655–659 (2005).

OBITUARY

Frank Albert Cotton (1930–2007)

Inorganic chemist, educator and discoverer of the quadruple bond.

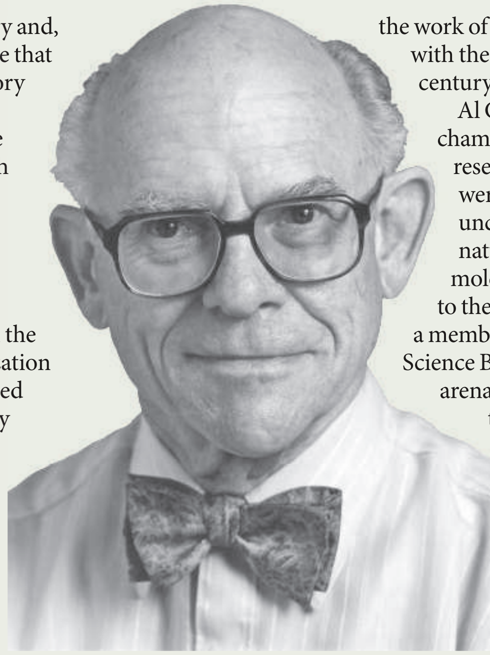
Among the greatest thrills for a scientist are those that occur at the moment of discovery. For a chemist, discovery is often synonymous with the creation of a new molecule in the laboratory. But understanding can be an even higher calling: the ability to recognize and explain properties of matter that others have previously encountered but could not comprehend. From understanding, one can derive new principles with which to interpret, or even predict, future results. Few individuals have contributed as much to inorganic chemistry through both discovery and understanding as Frank Albert Cotton. His death on 20 February, at the age of 76, marks the end of an extraordinary and prolific career.

The signature discovery to emerge from Cotton's laboratory at the Massachusetts Institute of Technology (MIT) came in the early 1960s. This was a time of renaissance in coordination chemistry — the study of metal–ligand complexes — and it was fuelled by the rediscovery of a theoretical model known as crystal field theory. Cotton and his colleagues applied this theory in their search for transition-metal complexes with unusual spin states and magnetic properties. They obtained rhenium compounds of relatively low oxidation state in which the distances between adjacent metal atoms were shorter than in the bulk metal itself.

The geometry of the simple $[Re_2Cl_8]^{2-}$ ion was especially noteworthy. In this ion, an extremely short metal–metal bond of just 2.27 Å linked two $\{ReCl_4^-\}$ fragments, with no additional bridging ligand. Moreover, the two halves of the dimer were in an 'eclipsed' configuration, with the chlorine atoms on either side swept back to a separation of about 3.5 Å to avoid non-bonded repulsions between them. For some reason, the unit adopted this geometry rather than avoiding such a steric clash by rotating one half of the dimer by 45°.

Cotton quickly understood the importance of this result. It meant that there was a quadruple bond in the $[Re_2Cl_8]^{2-}$ ion, the first to be identified. In addition to one σ and two π bonds between the rhenium atoms, as occurs in the triple bond between two nitrogen atoms in the N≡N molecule, there was an additional linkage, known as a δ bond. Formation of this bond required the chlorine atoms to adopt the eclipsed geometry.

Like all ground-breaking ideas in science, the consequences of this insight were profound. The finding that chemical compounds could have such multiply bonded metal–metal units launched a whole branch of inorganic chemistry and, later, materials science that kept Cotton's laboratory and others across the world occupied in the ensuing decades. Both experimentalists and theorists recognized the importance of this rich new vein of chemistry.



TEXAS A&M UNIV. DEPT CHEM.

An essential tool in the structural characterization of metal–metal-bonded compounds was X-ray crystallography. Cotton applied this technique so prodigiously, and advocated its more general use in structure determination so effectively, that it became a routine tool for investigating any chemical substance for which suitable single crystals could be obtained. With remarkable energy and creativity, the Cotton laboratory turned out a stunning display of compounds with double, triple and quadruple metal–metal bonds spanning many of the elements of the transition-metal series. By counting electrons, one could predict the character of many of these new creations — a remarkable achievement.

But Cotton's contributions were not limited to these metal-cluster compounds. In a series of imaginative studies, he and his colleagues prepared and investigated the physical and chemical properties of organometallic complexes. Many of these compounds contained carbon–carbon double bonds that were linked to one or more transition-metal ions. To clarify their structures, Cotton invented the 'hapto' or η nomenclature now used to indicate the number of carbon atoms bonded to a given metal atom.

Using nuclear magnetic resonance spectroscopy, typically applied to solutions of compounds over a temperature range, Cotton and his co-workers mapped out the relative motions of 'fluxional' organometallics, in which metal ions and organic fragments are involved in an elaborate intramolecular dance. They thus created fascinating motion pictures of movements within these complexes, and their studies were soon followed up by many laboratories. If a scientist's achievements can be measured by his or her ability to inspire the work of others, Cotton ranks with the very best twentieth-century chemists.

Al Cotton was a champion of fundamental research. His passions were learning, uncovering the secrets of nature and creating new molecules. As an adviser to the US government while a member of the National Science Board, and in other arenas, he championed the cause of funding basic research that is driven purely by intellectual curiosity, rather than by the latest fads in applying chemistry to its many sister disciplines in the life sciences and in engineering.

Cotton was also an extraordinary educator, writing textbooks at levels ranging from high school to advanced graduate, on topics from elementary and advanced inorganic chemistry to the application of group theory. These texts have been translated into many languages, and in total have sold more than a million copies. At MIT, and subsequently at Texas A&M University, he mentored well over 100 PhD students, and his laboratory was a popular training-ground for numerous postdocs and sabbatical visitors representing 30 countries.

Al took a deep personal interest in his postdoctoral and PhD students. I myself recall his wise advice about my future career choices when, as I was nearing the end of my graduate days in his lab at MIT, we were riding on horseback one beautiful autumn day near his home southwest of Boston. At that moment, and at many others spanning more than four decades, I was the beneficiary of his counsel and support, as were many others. His life in itself was an inspiration that taught the benefits and joys that reward hard work, deep thinking and the search for unexplored frontiers. ■

Stephen J. Lippard

Stephen J. Lippard is in the Department of Chemistry, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139-4307, USA.
e-mail: lippard@mit.edu

PALAEONTOLOGY

Undressing and redressing Ediacaran embryos

Arising from: J. V. Bailey, S. B. Joye, K. M. Kalanetra, B. E. Flood & F. A. Corsetti *Nature* **445,** 198–201 (2007)

Bailey *et al.*[1] propose that the Ediacaran microfossils *Megasphaera* and *Parapandorina*, previously interpreted as animal resting eggs and blastula embryos[2], represent *Thiomargarita*-like sulphide-oxidizing bacteria, claiming that this interpretation better explains their abundance and taphonomy. Here we highlight important observations that significantly weaken the authors' conclusions.

The first observations concern the morphological features of the microfossils. Morphotaxa should not be merged before their ontogeny and taphonomy are completely understood, and eggs and embryos were initially described[3] from the Ediacaran Doushantuo Formation in several taxa — *Megasphaera inornata* (one cell enclosed in a smooth envelope), *M. ornata* (one cell in an ornamented envelope), *Parapandorina raphospissa* (multiple polyhedral cells in a smooth envelope), and *Megaclonophycus onustus* (a large number of spheroidal cells in a smooth envelope). Subsequently, it was shown[4–6] that *M. ornata*, *Parapandorina* and a spiral microfossil (interpreted as a post-blastula embryo) are all surrounded by one or more envelopes.

The outermost envelope is typically ornamented with tubercular, polygonal, cerebral and fractal sculptures (Fig. 1a–i), and may also bear cylindrical processes[7]. The sculptured envelope is distinct from the botryoidal coating found on the cell surface or membrane of Doushantuo microfossils[5], and from the simple sheath of *Thiomargarita* or other bacteria[1]. However, it is similar to modern animal egg cases[3] and implies that there is a diapause stage — a physiological feature unknown in *Thiomargarita*.

Bailey *et al.*[1] propose that *M. inornata* and smooth-walled *Parapandorina* are *Thiomargarita*-like sulphur bacteria that underwent reductive cell division when stressed. However, their similar size, cell configuration, developmental sequence and occurrence indicate that the smooth-walled and ornamented populations may have close phylogenetic relationships. Alternatively, and more likely, the partial preservation of an ornamented envelope (Fig. 1a–g) at various cell-division stages indicates that these smooth-walled microfossils[1] are poorly preserved specimens whose ornamented outer envelope has been removed taphonomically[7]. Hence, the description by Bailey *et al.* of *Megasphaera* and *Parapandorina* is incomplete and their comparison to *Thiomargarita* inconclusive, so their interpretation is questionable.

Their bacterial interpretation[1] is also inconsistent with the organelle-like subcellular structures[8]. These structures are consistent in their size, shape, occurrence and location, indicating that their origin is biological rather than

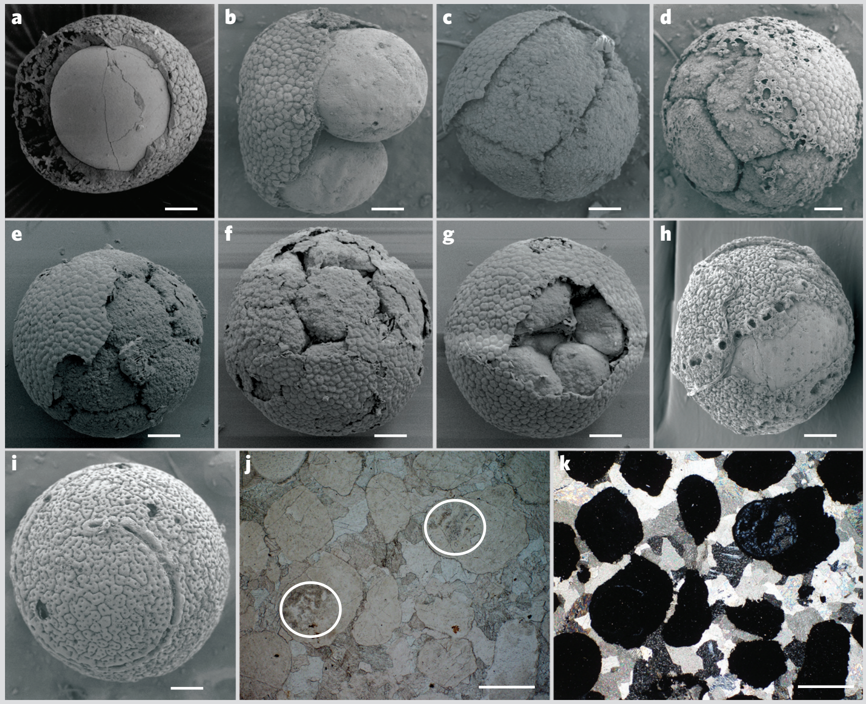


Figure 1 | Doushantuo microfossils. a–i, Scanning electron micoscope images showing partial preservation of the ornamented envelope in *Megasphaera ornata* (**a**), *Parapandorina raphospissa* (**b–g**) and a spiral microfossil (**h, i**)[6]. The apparently uneven thickness of the envelope in **a** is due to the eccentric location of the cell and to the unequal filling by cement in the space between the cell and its envelope. **j, k,** Plane-polarized (**j**) and cross-polarized (**k**) light micrographs of microfossils (circled in **j**) inside two phosphatic intraclasts in grey facies. Micrite is absent and intraclasts are cemented by dolospars. Scale bars: **a–i,** 100 μm; **j, k,** 500 μm.

taphonomic[8]. They differ from the irregularly shaped, degraded cytoplasmic structures in *Thiomargarita*[1]. Moreover, the spatial distribution of vacuoles and the cell numbers (typically fewer than eight) indicate that *Thiomargarita* is not really suitable for helping to interpret the Doushantuo microfossils[9].

Second, other observations on the taphonomic features of the microfossils undermine the conclusions of Bailey *et al.*[1]. The authors argue that a bacterial interpretation can explain how Doushantuo microfossils were phosphatized, because *Thiomargarita* concentrates phosphate in pore water. But if this were the case, why is phosphatized *Thiomargarita* not found in in modern sediments? Certainly, phosphate enrichment by some sort of bacteria may be important[10], but this may not mean that the phosphate-concentrating bacteria were exclusively or preferentially fossilized or self-entombed.

Bailey *et al.*[1] claim that the slow cell division of *Thiomargarita* allows a longer window for the preservation of the two-, four- and eight-cell stages of *Parapandorina*. The preservability of a *Parapandorina* individual is determined by the competing degradation and phosphatization processes, however, not by cell-division rate, because phosphatization occurs after death, not between cell divisions. The relative abundance of different cell-division stages in the preserved population reflects the age structure of the death assemblage, which may include individuals at any cell-division stage, irrespective of whether they are separated ontologically by years or minutes.

Bailey *et al.*[1] question whether Doushantuo microfossils were concentrated by reworking and winnowing[10,11], arguing that a bacterial interpretation explains their abundance. *Thiomargarita* cells are abundant in modern sediments, and *Megasphaera* and *Parapandorina* are abundant in the grey facies of Doushantuo phosphorites[11]. However, abrasion, rounding and inclusion in intraclasts[12] (Fig. 1j, k) is indicative of intrabasinal reworking and winnowing, which concentrated Doushantuo microfossils in the grey facies. *Megasphaera* and *Parapandorina* also occur in the less-reworked black facies of Doushantuo phosphorites[11], but are much less abundant. In fact, there are so few in the more pristine black facies that a comparison with the abundant occurrence of *Thiomargarita* seems inappropriate.

The naked Doushantuo microfossils illustrated by Bailey *et al.*[1] were therefore probably undressed by taphonomic processes; when redressed and considered alongside other Doushantuo fossils, their similarity to *Thiomargarita* is restricted to reductive cell division — a convergent response to a limited external nutrient supply. In our view, their bacterial interpretation does not adequately address the morphology, taphonomy and abundance of Doushantuo microfossils.

Shuhai Xiao*, Chuanming Zhou†, Xunlai Yuan†
*Department of Geosciences, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24060, USA
e-mail: xiao@vt.edu
†State Key Laboratory of Paleobiology and Stratigraphy, Nanjing Institute of Geology and Paleontology, Chinese Academy of Sciences, Nanjing 210008, China

Received 7 January 2007; accepted 20 February 2007.

1. Bailey, J. V., Joye, S. B., Kalanetra, K. M., Flood, B. E. & Corsetti, F. A. *Nature* **445,** 198–201 (2007); published onlne 20 December 2006, doi:10.1038/nature05457.
2. Xiao, S., Zhang, Y. & Knoll, A. H. *Nature* **391,** 553–558 (1998).
3. Xiao, S. & Knoll, A. H. *J. Paleontol.* **74,** 767–788 (2000).
4. Yuan, X. *et al. Doushantuo Fossils: Life on the Eve of Animal Radiation* (China Univ. of Science and Technology Press, Hefei, China, 2002).
5. Chen, J.-Y. *The Dawn of Animal World* (Jiangsu Science and Technology, Nanjing, 2004).
6. Xiao, S., Hagadorn, J. W., Zhou, C. & Yuan, X. *Geology* **35,** 115–118 (2007).
7. Yin, C., Bengtson, S. & Yue, Z. *Acta Palaeontol. Polon.* **49,** 1–12 (2004).
8. Hagadorn, J. W. *et al. Science* **314,** 291–294 (2006).
9. Donoghue, P. C. J. *Nature* **445,** 155–156 (2007).
10. Xiao, S. & Knoll, A. H. *Lethaia* **32,** 219–240 (1999).
11. Dornbos, S. Q. *et al. Palaios* **21,** 3–14 (2006).
12. Zhang, Y., Yin, L., Xiao, S. & Knoll, A. H. *Paleontol. Soc. Mem.* **50,** 1–52 (1998).

Competing financial interests: declared none.
doi:10.1038/nature05753

PALAEONTOLOGY

Bailey et al. *reply*

Replying to: S. Xiao, C. Zhou & X. Yuan *Nature* **446,** doi: 10.1038/nature05753 (2007)

Xiao *et al.*[1] suggest that the presence of a textured capsule surrounding some Doushantuo globular microfossils calls into question the alternative interpretation of these structures as giant sulphur bacteria similar to modern *Thiomargarita*[2]. However, the outer coatings illustrated by Xiao *et al.*[1] are similar morphologically to known bacterial features, and the texture, location and thickness change of the capsule is inconsistent with that of a fertilization envelope. We are therefore not convinced that the bacterial hypothesis has been falsified.

Xiao *et al.*[1] provide new images of the Doushantuo globular structures that reveal a thick, ornamented outer capsule on some specimens, and suggest that the patterns on the capsule are too complex to be microbial in origin. Our original discussion centred on Doushantuo specimens that had smooth envelopes (*Megasphaera inornata* and *Parapandorina*), as smooth and textured examples had been assigned different linnaean names[3] — the implication being that they were different organisms. On the one hand, we note that the embryo and bacterial hypotheses are not mutually exclusive: both could have existed together in the same deposit, just as eukaryotic debris can be found with *Thiomargarita* today. This possibility might explain the abundance of specimens that have smooth (or entirely lack) envelopes, the absence of *Tianzhushania*-like peripheral processes in ubiquitous phosphatized specimens (but see ref. 4) and the multiple geometric configurations seen in cell clusters at the same stage of reductive division, which are not easily explained by a single metazoan species. On the other hand, we welcome the opportunity to discuss the textured examples in light of the bacterial hypothesis, given the tacit assumption by Xiao *et al.* that all Doushantuo globular structures have at one time had such elaborate envelopes.

Thiomargarita from the Gulf of Mexico are surrounded by a multilayered ultrastructure and display a mucus-filled sheath[2], similar in thickness to the capsule noted by Xiao *et al.*, which, if fossilized, would make an excellent textural match to the Doushantuo envelopes. As for the complex pattern illustrated by Xiao *et al.*, similar complex patterns have been well documented from the microbial world[5–8]. Inclusion-bearing sulphide-oxidizing bacteria in the same family as *Thiomargarita* (Fig. 1a) have strikingly similar outer textures[8] to those illustrated by Xiao *et al.*, demonstrating that such features are not exclusive to the Eukarya. Thus, the presence of an ornamented capsule is consistent with our hypothesis.

Note that the capsule surrounding *Megasphaera ornata* shown by Xiao *et al.*[1] exhibits an extreme variation in thickness across the structure (Fig. 1b). A sculptured capsule is also seen to enclose two internal bodies on their distal sides and to act as a medial boundary between the same two bodies (Fig. 1c). Such occurrences are incompatible with the interpretation of the textured capsule as a fertilization envelope. Additionally, the spiral grooves and pits on capsules shown by Xiao *et al.*[1] show a remarkably similar geometry to pyrite trails[9] considered to be abiotic[10]. These features therefore do not provide a valid falsification of the bacterial hypothesis.

Xiao *et al.* are also compelled by the presence of paired reniform structures noted by computed-tomography techniques within a few *Parapandorina* specimens[11], and suggest that their presence falsifies the bacterial hypothesis. We disagree. Only ten (of 162, screened from thousands) contain large internal structures, and only a few of those would be characterized as 'regular' in appearance. The reniform sulphur- or polyphosphate-inclusion clusters seen in degrading *Thiomargarita*[2] may better explain the occurrence of compositionally distinct regions that occur in their specimens.

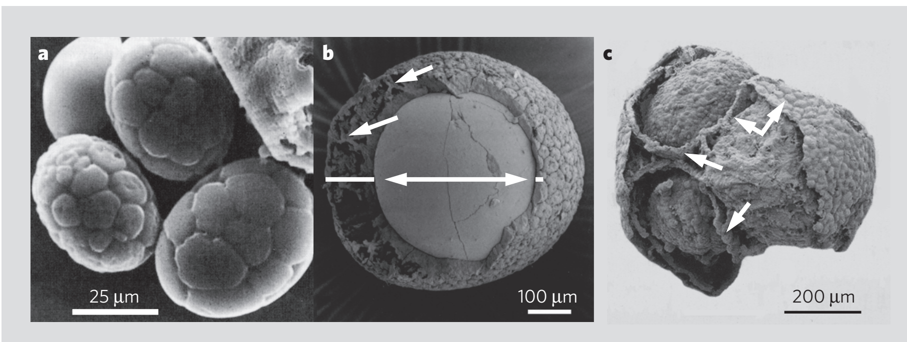


Figure 1 | Interpretation of microfossil bodies enclosed in textured envelopes. **a**, **b**, An exclusively metazoan interpretation is called into question by **a**, complex patterns on bacterial cells, such as polygonal patterns on inclusion-bearing *Achromatium*, a close relative of *Thiomargarita*[8]; by diagenetic features in Doushantuo microfossils such as **b**, *Megasphaera ornata*, a specimen showing internal diagenetic processes (single arrows) and thickening of the envelope[3] (double arrows and lines), and **c**, *Parapandorina*, a specimen showing an ornamented envelope that passes between internal bodies[10].

Features once considered diagnostic for one group of microbes (such as size, cell-division and complex ultrastructural patterns) are now found in other groups initially thought to be too simple for such features. Perhaps a more convincing test of the bacterial and animal interpretations will arise in the future. For now, we are content to say the emperor has no clothes with respect to the undressing and redressing of the Doushantuo globular structures.

Jake V. Bailey*, Samantha B. Joye‡, Karen M. Kalanetra‡, Beverly E. Flood†, Frank A. Corsetti*

*Department of Earth Sciences and †Department of Biological Sciences, University of Southern California, Los Angeles, California 90089, USA
e-mail: jvbailey@usc.edu
‡Department of Marine Sciences, University of Georgia, Athens, Georgia 30602, USA

1. Xiao, S., Zhou, C. & Yuan, X. *Nature* **446,** doi: 10.1038/nature05753 (2007).
2. Bailey, J. V., Joye, S. B., Kalenetra, K. M., Flood, B. E. & Corsetti, F. A. *Nature* **445,** 198–201 (2007); published onlne 20 December 2006, doi:10.1038/nature05457.
3. Xiao, S. & Knoll, A. H. *J. Paleontol.* **74,** 767–788 (2000).
4. Yin, C., Bengston, S. & Yue, Z. *Acta Palaeontol. Polon.* **49,** 1–12 (2004).
5. Strohl, W. R., Howard, K. S. & Larkin, J. M. *J. Gen. Microbiol.* **128,** 73–84 (1982).
6. Thornley, M. J., Glauert, A. M. & Sleytr, U. B. *J. Bacteriol.* **114**, 1294–1308 (1973).
7. Thar, R. & Kühl, M. *Appl. Envir. Microbiol.* **68,** 6310–6320 (2002).
8. Babenzien, H. D. & Sass, H. in *Ergebnisse der Limnologie: Aquatic Microbial Ecology* (eds Simon, M., Güde, H. & Weisse, T.) Vol. 48, 247–251 (Schweizerbart, Stuttgart, 1996).
9. Chen, J.-Y. *The Dawn of Animal World* (Jiangsu Science and Technology, Nanjing, 2004).
10. Xiao, S. & Knoll, A. H. *Lethaia* **32,** 219–240 (1999).
11. Hagadorn, J. W. *et al. Science* **314,** 291–294 (2006).

doi:10.1038nature/05754

PALAEONTOLOGY

Undressing and redressing Ediacaran embryos

Arising from: J. V. Bailey, S. B. Joye, K. M. Kalanetra, B. E. Flood & F. A. Corsetti *Nature* **445,** 198–201 (2007)

Bailey *et al.*[1] propose that the Ediacaran microfossils *Megasphaera* and *Parapandorina*, previously interpreted as animal resting eggs and blastula embryos[2], represent *Thiomargarita*-like sulphide-oxidizing bacteria, claiming that this interpretation better explains their abundance and taphonomy. Here we highlight important observations that significantly weaken the authors' conclusions.

The first observations concern the morphological features of the microfossils. Morphotaxa should not be merged before their ontogeny and taphonomy are completely understood, and eggs and embryos were initially described[3] from the Ediacaran Doushantuo Formation in several taxa — *Megasphaera inornata* (one cell enclosed in a smooth envelope), *M. ornata* (one cell in an ornamented envelope), *Parapandorina raphospissa* (multiple polyhedral cells in a smooth envelope), and *Megaclonophycus onustus* (a large number of spheroidal cells in a smooth envelope). Subsequently, it was shown[4–6] that *M. ornata*, *Parapandorina* and a spiral microfossil (interpreted as a post-blastula embryo) are all surrounded by one or more envelopes.

The outermost envelope is typically ornamented with tubercular, polygonal, cerebral and fractal sculptures (Fig. 1a–i), and may also bear cylindrical processes[7]. The sculptured envelope is distinct from the botryoidal coating found on the cell surface or membrane of Doushantuo microfossils[5], and from the simple sheath of *Thiomargarita* or other bacteria[1]. However, it is similar to modern animal egg cases[3] and implies that there is a diapause stage — a physiological feature unknown in *Thiomargarita*.

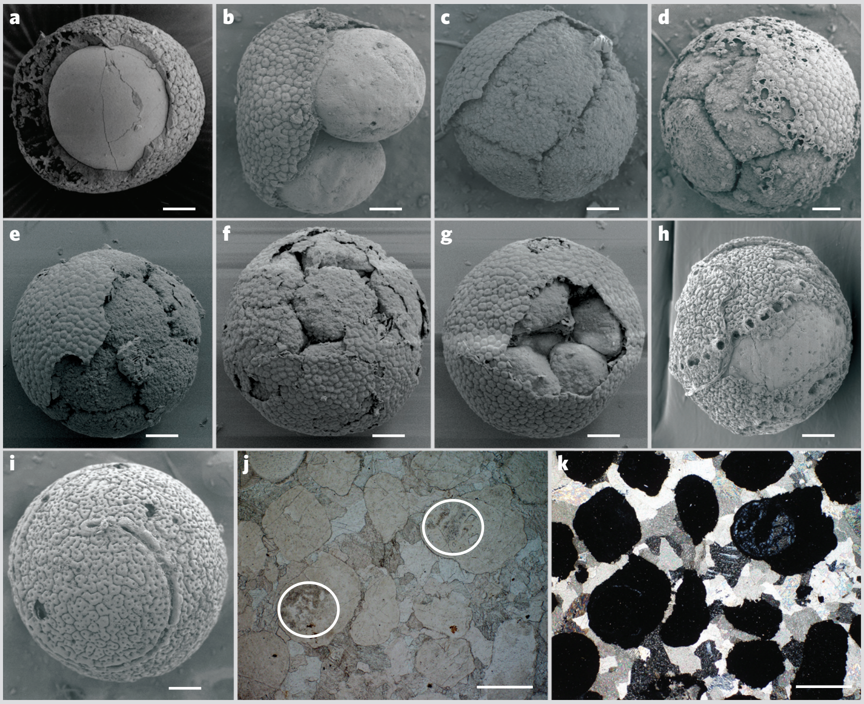


Figure 1 | Doushantuo microfossils. a–i, Scanning electron micoscope images showing partial preservation of the ornamented envelope in *Megasphaera ornata* (**a**), *Parapandorina raphospissa* (**b–g**) and a spiral microfossil (**h, i**)[6]. The apparently uneven thickness of the envelope in **a** is due to the eccentric location of the cell and to the unequal filling by cement in the space between the cell and its envelope. **j, k,** Plane-polarized (**j**) and cross-polarized (**k**) light micrographs of microfossils (circled in **j**) inside two phosphatic intraclasts in grey facies. Micrite is absent and intraclasts are cemented by dolospars. Scale bars: **a–i**, 100 μm; **j, k**, 500 μm.

Bailey *et al.*[1] propose that *M. inornata* and smooth-walled *Parapandorina* are *Thiomargarita*-like sulphur bacteria that underwent reductive cell division when stressed. However, their similar size, cell configuration, developmental sequence and occurrence indicate that the smooth-walled and ornamented populations may have close phylogenetic relationships. Alternatively, and more likely, the partial preservation of an ornamented envelope (Fig. 1a–g) at various cell-division stages indicates that these smooth-walled microfossils[1] are poorly preserved specimens whose ornamented outer envelope has been removed taphonomically[7]. Hence, the description by Bailey *et al.* of *Megasphaera* and *Parapandorina* is incomplete and their comparison to *Thiomargarita* inconclusive, so their interpretation is questionable.

Their bacterial interpretation[1] is also inconsistent with the organelle-like subcellular structures[8]. These structures are consistent in their size, shape, occurrence and location, indicating that their origin is biological rather than taphonomic[8]. They differ from the irregularly shaped, degraded cytoplasmic structures in *Thiomargarita*[1]. Moreover, the spatial distribution of vacuoles and the cell numbers (typically fewer than eight) indicate that *Thiomargarita* is not really suitable for helping to interpret the Doushantuo microfossils[9].

Second, other observations on the taphonomic features of the microfossils undermine the conclusions of Bailey *et al.*[1]. The authors argue that a bacterial interpretation can explain how Doushantuo microfossils were phosphatized, because *Thiomargarita* concentrates phosphate in pore water. But if this were the case, why is phosphatized *Thiomargarita* not found in in modern sediments? Certainly, phosphate enrichment by some sort of bacteria may be important[10], but this may not mean that the phosphate-concentrating bacteria were exclusively or preferentially fossilized or self-entombed.

Bailey *et al.*[1] claim that the slow cell division of *Thiomargarita* allows a longer window for the preservation of the two-, four- and eight-cell stages of *Parapandorina*. The preservability of a *Parapandorina* individual is determined by the competing degradation and phosphatization processes, however, not by cell-division rate, because phosphatization occurs after death, not between cell divisions. The relative abundance of different cell-division stages in the preserved population reflects the age structure of the death assemblage, which may include individuals at any cell-division stage, irrespective of whether they are separated ontologically by years or minutes.

Bailey *et al.*[1] question whether Doushantuo microfossils were concentrated by reworking and winnowing[10,11], arguing that a bacterial interpretation explains their abundance. *Thiomargarita* cells are abundant in modern sediments, and *Megasphaera* and *Parapandorina* are abundant in the grey facies of Doushantuo phosphorites[11]. However, abrasion, rounding and inclusion in intraclasts[12] (Fig. 1j, k) is indicative of intrabasinal reworking and winnowing, which concentrated Doushantuo microfossils in the grey facies. *Megasphaera* and *Parapandorina* also occur in the less-reworked black facies of Doushantuo phosphorites[11], but are much less abundant. In fact, there are so few in the more pristine black facies that a comparison with the abundant occurrence of *Thiomargarita* seems inappropriate.

The naked Doushantuo microfossils illustrated by Bailey *et al.*[1] were therefore probably undressed by taphonomic processes; when redressed and considered alongside other Doushantuo fossils, their similarity to *Thiomargarita* is restricted to reductive cell division — a convergent response to a limited external nutrient supply. In our view, their bacterial interpretation does not adequately address the morphology, taphonomy and abundance of Doushantuo microfossils.

Shuhai Xiao*, Chuanming Zhou†, Xunlai Yuan†
*Department of Geosciences, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24060, USA
e-mail: xiao@vt.edu
†State Key Laboratory of Paleobiology and Stratigraphy, Nanjing Institute of Geology and Paleontology, Chinese Academy of Sciences, Nanjing 210008, China

Received 7 January 2007; accepted 20 February 2007.

1. Bailey, J. V., Joye, S. B., Kalanetra, K. M., Flood, B. E. & Corsetti, F. A. *Nature* **445,** 198–201 (2007); published onlne 20 December 2006, doi:10.1038/nature05457.
2. Xiao, S., Zhang, Y. & Knoll, A. H. *Nature* **391,** 553–558 (1998).
3. Xiao, S. & Knoll, A. H. *J. Paleontol.* **74,** 767–788 (2000).
4. Yuan, X. *et al. Doushantuo Fossils: Life on the Eve of Animal Radiation* (China Univ. of Science and Technology Press, Hefei, China, 2002).
5. Chen, J.-Y. *The Dawn of Animal World* (Jiangsu Science and Technology, Nanjing, 2004).
6. Xiao, S., Hagadorn, J. W., Zhou, C. & Yuan, X. *Geology* **35,** 115–118 (2007).
7. Yin, C., Bengtson, S. & Yue, Z. *Acta Palaeontol. Polon.* **49,** 1–12 (2004).
8. Hagadorn, J. W. *et al. Science* **314,** 291–294 (2006).
9. Donoghue, P. C. J. *Nature* **445,** 155–156 (2007).
10. Xiao, S. & Knoll, A. H. *Lethaia* **32,** 219–240 (1999).
11. Dornbos, S. Q. *et al. Palaios* **21,** 3–14 (2006).
12. Zhang, Y., Yin, L., Xiao, S. & Knoll, A. H. *Paleontol. Soc. Mem.* **50,** 1–52 (1998).

Competing financial interests: declared none.
doi:10.1038/nature05753

PALAEONTOLOGY

Bailey et al. *reply*

Replying to: S. Xiao, C. Zhou & X. Yuan *Nature* **446,** doi: 10.1038/nature05753 (2007)

Xiao *et al.*[1] suggest that the presence of a textured capsule surrounding some Doushantuo globular microfossils calls into question the alternative interpretation of these structures as giant sulphur bacteria similar to modern *Thiomargarita*[2]. However, the outer coatings illustrated by Xiao *et al.*[1] are similar morphologically to known bacterial features, and the texture, location and thickness change of the capsule is inconsistent with that of a fertilization envelope. We are therefore not convinced that the bacterial hypothesis has been falsified.

Xiao *et al.*[1] provide new images of the Doushantuo globular structures that reveal a thick, ornamented outer capsule on some specimens, and suggest that the patterns on the capsule are too complex to be microbial in origin. Our original discussion centred on Doushantuo specimens that had smooth envelopes (*Megasphaera inornata* and *Parapandorina*), as smooth and textured examples had been assigned different linnaean names[3] — the implication being that they were different organisms. On the one hand, we note that the embryo and bacterial hypotheses are not mutually exclusive: both could have existed together in the same deposit, just as eukaryotic debris can be found with *Thiomargarita* today. This possibility might explain the abundance of specimens that have smooth (or entirely lack) envelopes, the absence of *Tianzhushania*-like peripheral processes in ubiquitous phosphatized specimens (but see ref. 4) and the multiple geometric configurations seen in cell clusters at the same stage of reductive division, which are not easily explained by a single metazoan species. On the other hand, we welcome the opportunity to discuss the textured examples in light of the bacterial hypothesis, given the tacit assumption by Xiao *et al.* that all Doushantuo globular structures have at one time had such elaborate envelopes.

Thiomargarita from the Gulf of Mexico are surrounded by a multilayered ultrastructure and display a mucus-filled sheath[2], similar in thickness to the capsule noted by Xiao *et al.*, which, if fossilized, would make an excellent textural match to the Doushantuo envelopes. As for the complex pattern illustrated by Xiao *et al.*, similar complex patterns have been well documented from the microbial world[5–8]. Inclusion-bearing sulphide-oxidizing bacteria in the same family as *Thiomargarita* (Fig. 1a) have strikingly similar outer textures[8] to those illustrated by Xiao *et al.*, demonstrating that such features are not exclusive to the Eukarya. Thus, the presence of an ornamented capsule is consistent with our hypothesis.

Note that the capsule surrounding *Megasphaera ornata* shown by Xiao *et al.*[1] exhibits an extreme variation in thickness across the structure (Fig. 1b). A sculptured capsule is also seen to enclose two internal bodies on their distal sides and to act as a medial boundary between the same two bodies (Fig. 1c). Such occurrences are incompatible with the interpretation of the textured capsule as a fertilization envelope. Additionally, the spiral grooves and pits on capsules shown by Xiao *et al.*[1] show a remarkably similar geometry to pyrite trails[9] considered to be abiotic[10]. These features therefore do not provide a valid falsification of the bacterial hypothesis.

Xiao *et al.* are also compelled by the presence of paired reniform structures noted by computed-tomography techniques within a few *Parapandorina* specimens[11], and suggest that their presence falsifies the bacterial hypothesis. We disagree. Only ten (of 162, screened from thousands) contain large internal structures, and only a few of those would be characterized as 'regular' in appearance. The reniform sulphur- or polyphosphate-inclusion clusters seen in degrading *Thiomargarita*[2] may better explain the occurrence of compositionally distinct regions that occur in their specimens.

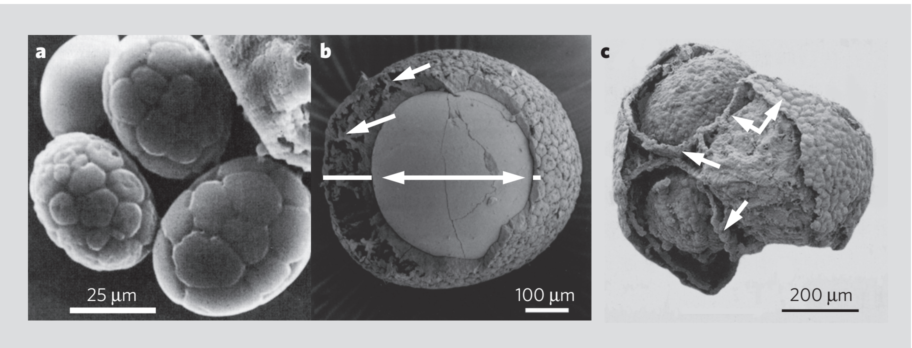


Figure 1 | Interpretation of microfossil bodies enclosed in textured envelopes. a, b, An exclusively metazoan interpretation is called into question by **a**, complex patterns on bacterial cells, such as polygonal patterns on inclusion-bearing *Achromatium*, a close relative of *Thiomargarita*[8]; by diagenetic features in Doushantuo microfossils such as **b**, *Megasphaera ornata*, a specimen showing internal diagenetic processes (single arrows) and thickening of the envelope[3] (double arrows and lines), and **c**, *Parapandorina*, a specimen showing an ornamented envelope that passes between internal bodies[10].

Features once considered diagnostic for one group of microbes (such as size, cell-division and complex ultrastructural patterns) are now found in other groups initially thought to be too simple for such features. Perhaps a more convincing test of the bacterial and animal interpretations will arise in the future. For now, we are content to say the emperor has no clothes with respect to the undressing and redressing of the Doushantuo globular structures.

Jake V. Bailey*, Samantha B. Joye‡, Karen M. Kalanetra‡, Beverly E. Flood†, Frank A. Corsetti*

*Department of Earth Sciences and †Department of Biological Sciences, University of Southern California, Los Angeles, California 90089, USA
e-mail: jvbailey@usc.edu
‡Department of Marine Sciences, University of Georgia, Athens, Georgia 30602, USA

1. Xiao, S., Zhou, C. & Yuan, X. *Nature* **446,** doi: 10.1038/nature05753 (2007).
2. Bailey, J. V., Joye, S. B., Kalenetra, K. M., Flood, B. E. & Corsetti, F. A. *Nature* **445,** 198–201 (2007); published onlne 20 December 2006, doi:10.1038/nature05457.
3. Xiao, S. & Knoll, A. H. *J. Paleontol.* **74,** 767–788 (2000).
4. Yin, C., Bengston, S. & Yue, Z. *Acta Palaeontol. Polon.* **49,** 1–12 (2004).
5. Strohl, W. R., Howard, K. S. & Larkin, J. M. *J. Gen. Microbiol.* **128,** 73–84 (1982).
6. Thornley, M. J., Glauert, A. M. & Sleytr, U. B. *J. Bacteriol.* **114**, 1294–1308 (1973).
7. Thar, R. & Kühl, M. *Appl. Envir. Microbiol.* **68,** 6310–6320 (2002).
8. Babenzien, H. D. & Sass, H. in *Ergebnisse der Limnologie: Aquatic Microbial Ecology* (eds Simon, M., Güde, H. & Weisse, T.) Vol. 48, 247–251 (Schweizerbart, Stuttgart, 1996).
9. Chen, J.-Y. *The Dawn of Animal World* (Jiangsu Science and Technology, Nanjing, 2004).
10. Xiao, S. & Knoll, A. H. *Lethaia* **32,** 219–240 (1999).
11. Hagadorn, J. W. *et al. Science* **314,** 291–294 (2006).

doi:10.1038nature/05754

Vol 446 | 5 April 2007 | doi:10.1038/nature05648

ARTICLES

Attosecond real-time observation of electron tunnelling in atoms

M. Uiberacker[1,2], Th. Uphues[3], M. Schultze[2], A. J. Verhoef[2,4], V. Yakovlev[1], M. F. Kling[5], J. Rauschenberger[1,2], N. M. Kabachnik[3,6], H. Schröder[2], M. Lezius[2], K. L. Kompa[2], H.-G. Muller[5], M. J. J. Vrakking[5], S. Hendel[3], U. Kleineberg[1], U. Heinzmann[3], M. Drescher[7] & F. Krausz[1,2,4]

Atoms exposed to intense light lose one or more electrons and become ions. In strong fields, the process is predicted to occur via tunnelling through the binding potential that is suppressed by the light field near the peaks of its oscillations. Here we report the real-time observation of this most elementary step in strong-field interactions: light-induced electron tunnelling. The process is found to deplete atomic bound states in sharp steps lasting several hundred attoseconds. This suggests a new technique, attosecond tunnelling, for probing short-lived, transient states of atoms or molecules with high temporal resolution. The utility of attosecond tunnelling is demonstrated by capturing multi-electron excitation (shake-up) and relaxation (cascaded Auger decay) processes with subfemtosecond resolution.

With the invention of the laser in 1960 and subsequent advances in ultrashort light pulse generation, light fields with a strength of several volts per angstrom have become routinely available. They rival the fields acting on valence electrons in atomic systems, allowing their release from atoms, molecules and solids. These advances sparked a revolution in studying the interaction of electrons with light. The primary step in strong-field interactions is the liberation of electrons from their atomic bound state. The revolutionary theory of Keldysh[1] and subsequent work[2–6] suggested that a valence electron may escape by tunnelling through its atomic binding potential suppressed by the light field (Fig. 1a). If the dimensionless parameter:

$$\gamma = \frac{\omega_{\mathrm{L}}\sqrt{2mW_{\mathrm{b}}}}{|e|E_0} \quad (1)$$

is less than one, under the assumption of $\hbar\omega_{\mathrm{L}} \ll W_{\mathrm{b}}$ ionization is predicted to be confined to short intervals lasting a fraction of the half oscillation cycle of the light field (Fig. 1b). Here E_0 and ω_{L} stand for the amplitude and angular frequency of the oscillations of the laser electric field $E_{\mathrm{L}}(t) = E_0\varepsilon(t)\cos(\omega_{\mathrm{L}}t + \varphi)$, with $\varepsilon(t)$ being the amplitude envelope function, and e, m and W_{b} the charge, mass and binding energy of the electron. Recent studies[6] suggest that tunnelling remains the dominant ionization mechanism even for γ substantially exceeding one, that is, under conditions when the potential barrier formed by the atomic binding potential and the ionizing light field varies during tunnelling (non-adiabatic regime).

In this work we report what we believe is the first real-time observation of light-induced electron tunnelling. The observation of ionization occurring in subfemtosecond steps spaced by the half laser cycle up to values of the Keldysh parameter as high as three is in good agreement with analytic and numerical calculations. Our approach provides experimental access to all forms of optical field ionization—both adiabatic and nonadiabatic tunnelling, as well as barrier-suppression ionization—and allows us to test models of these processes for the first time.

Once the process of field ionization is fully understood, the technique of attosecond tunnelling will provide direct time-domain insight into a wide range of multi-electron dynamics and electron–electron interactions, ultimately with a resolution approaching the atomic unit of time (~24 as). We demonstrate this potential by probing shake-up and Auger cascade processes with subfemtosecond resolution.

Attosecond probing of electron dynamics

Figure 2 illustrates different options for attosecond sampling of electronic motion in atoms or molecules. A subfemtosecond extreme ultraviolet (XUV) pulse triggers the motion by exciting a valence or core electron (Fig. 2a, b). The unfolding excitation and relaxation processes (Fig. 2) could, in principle, be probed by a delayed replica of the pulse. However, the low flux of currently available subfemtosecond XUV pulses and the low two-photon transition probabilities in the XUV and X-ray regimes have thwarted this straightforward extension of conventional pump–probe techniques into the XUV–X-ray spectral range.

A few-cycle wave of visible or near-infrared (NIR) light with controlled waveform[7] in combination with a highly nonlinear process may replace the subfemtosecond pulse either in probing or starting electron dynamics. This was first demonstrated by attosecond streaking[8,9]: the strong-field interaction of a few-cycle light wave with free electrons released by a subfemtosecond XUV excitation pulse results in broadening and shifting of their final momentum distribution. Recording the streaked spectra of the emitted photo- and Auger electrons versus delay between the XUV pump and the few-cycle laser probe allowed us to retrieve the XUV pulse[10,11] and the laser field[12] as well as the inner-atomic relaxation dynamics[13] with subfemtosecond resolution.

Here, we demonstrate that nonlinear interaction of the same light wave with bound electrons ionizes in subfemtosecond steps and hence offers a means of probing intra-atomic and intra-molecular electron dynamics—including when no free electrons are released—by means of attosecond tunnelling. This approach relies on the fact that energetic photo-excitation as well as subsequent rearrangement

[1]Department für Physik, Ludwig-Maximilians-Universität, Am Coulombwall 1, [2]Max-Planck-Institut für Quantenoptik, Hans-Kopfermann-Strasse 1, D-85748 Garching, Germany. [3]Fakultät für Physik, Universität Bielefeld, Universitätsstrasse 25, D-33615 Bielefeld, Germany. [4]Technische Universität Wien, Gusshausstrasse 27, A-1040 Vienna, Austria. [5]FOM-Instituut voor Atoom- en Molecuulfysica (AMOLF), Kruislaan 407, 1098 SJ, Amsterdam, The Netherlands. [6]Institute of Nuclear Physics, Moscow State University, Moscow 119992, Russia. [7]Institut für Experimentalphysik, Universität Hamburg, Luruper Chaussee 149, D-22671 Hamburg, Germany.

via Auger decay is accompanied by transitions to unoccupied orbitals via 'shake-up' (Fig. 2d–f)[14–16]. Here, we use the term 'shake-up' in a broad sense, standing for all possible processes populating excited ionic states (including instantaneous and non-instantaneous ones), henceforth referred to as shake-up states (represented by levels 1, 2 and 3 in the example of Fig. 2). These populations can be probed via optical field ionization by a strong, few-cycle NIR pulse of variable delay Δt, with respect to the subfemtosecond XUV excitation (Fig. 1c) by measuring the number of ions resulting from the XUV pump–NIR probe exposure as a function of Δt (Fig. 1d).

Shake-up usually populates several quantum states in the valence band, from which electrons can be freed by the NIR probe. Hence, the ion yield will constitute an integral signal, with contributions from all shake-up states up to a certain binding energy from which ionization is feasible for the intensity chosen. Population dynamics of individual states of significantly differing binding energy can be retrieved by pump–probe scans repeated at different NIR probe intensities and/or from the temporal separation of the depletion of the states in the same delay scan, as explained in Fig. 1d and demonstrated in Fig. 4. The ion yield constitutes an integral signal in a temporal sense, too. The shake-up states are exposed to the ionizing NIR field from the moment they have been populated until the end of the NIR pulse. The beginning of this time interval within the NIR pulse is adjusted with the delay Δt between the XUV pump and the NIR probe. As the delay is scanned from large negative values (NIR probe first) to large positive values (XUV pump first) the measured ion yield (Fig. 1d) starts increasing at $\Delta t < 0$ owing to ionization on the trailing edge of the NIR pulse and continues to increase with increasing Δt because the XUV pump shifts towards the peak of the ionizing NIR probe and the shake-up states are exposed to ever higher NIR probe intensities.

In the absence of Auger decay, shake-up excitation results from photoionization only (Fig. 2d). The time-dependent ionization dynamics sketched in Fig. 1d can then be traced by measuring the yield of doubly charged ions as a function of the delay between the XUV pump and the sampling NIR light field. The temporal ionization gradients sketched in Fig. 1d incorporate the finite duration of the XUV excitation, a possible delayed response of shake-up and the tunnelling dynamics. With a sufficiently rapid ($\ll$1 fs) excitation, these measurements can thus provide direct insight into the temporal evolution of shake-up (in the presence of a strong optical field) and light-induced tunnelling. In what follows, first we prove this concept by exposing neon atoms to our subfemtosecond XUV pump and few-cycle NIR probe pulses (Fig. 3) and measuring the yield of the product of the pump–probe exposure, Ne^{2+}, versus Δt (Fig. 4). Then we extend the approach to probing electrons shaken up in xenon atoms during Auger decay. Our primary observables in this case are higher charged states, Xe^{N+} (for $N > 2$). Measuring their yield versus Δt displays the course of an Auger cascade.

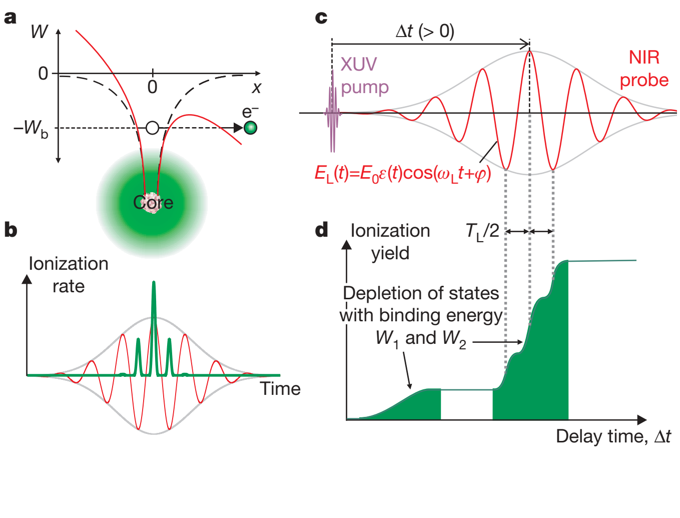


Figure 1 | Strong-field ionization and pump-probe setting for its real-time observation. **a**, Exposing an atom to a strong NIR laser field will result in a modified potential (solid curve) composed of the Coulomb potential (dashed curve) and the time-dependent effective potential of the laser pulse. The laser is polarized along the x direction and W_b is the binding energy of the electron. At sufficiently high laser field strengths the atomic binding potential is suppressed to a small barrier in the x or $-x$ direction for the maxima and minima of the laser electric field, respectively, allowing optical tunnelling to become the dominant ionization mechanism. **b**, The highly nonlinear dependence of the tunnelling rate on the width of the potential barrier confines ionization to time intervals of very short duration near the field oscillation maxima. In the electric field of a few-cycle NIR laser pulse (thin line) ionization is predicted to be restricted to several subfemtosecond intervals (thick line). **c**, Concept for tracing optical-field ionization: a subfemtosecond XUV pulse generates ions in excited (shake-up) states, from which a time-delayed NIR few-cycle probe pulse liberates electrons to produce doubly charged ions. A delay $\Delta t = 0$ is defined as the coincidence of the peaks of the envelopes of the NIR and XUV pulse. $\Delta t > 0$ implies that the peak of the XUV pulse envelope precedes that of the NIR pulse. **d**, Yield of doubly charged ions versus delay Δt between the XUV pump and the NIR probe, as predicted qualitatively for the case of electrons being prepared in two shake-up states of significantly differing binding energy (W_1 and $W_2 \gg W_1$) for liberation via strong-field ionization (see text for discussion). The state of low binding energy is depleted at low intensities, where multiphoton ionization dominates and hence sub-cycle structure in the ionization dynamics is absent. By contrast, the state of high binding energy is emptied by tunnelling, resulting in pronounced sub-half-cycle steps in the ionization profile.

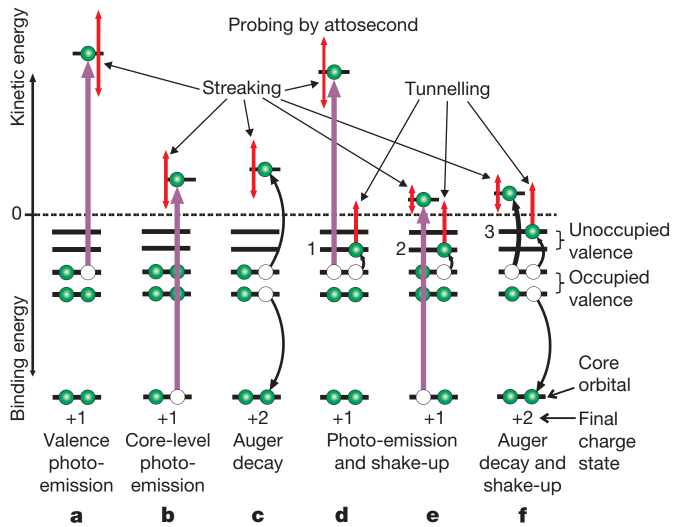


Figure 2 | Probing electron dynamics in atoms, molecules or solids with attosecond sampling techniques. A subfemtosecond XUV pulse triggers the motion by inducing valence (process **a**) or core photoelectron emission (process **b**). The temporal evolution of photo- and Auger electron emission (process **c**) can be probed via attosecond streaking to retrieve the XUV pump pulse or the sampling NIR field and trace inner-shell relaxation dynamics. XUV photoexcitation as well as subsequent Auger decay processes are usually accompanied by shake-up of another electron to a previously unoccupied level (processes **d**, **e** and **f**). In this case the liberated electrons will be ejected at a reduced kinetic energy compared to the cases without shake-up processes **a**, **b** and **c**. The difference in energy is used for shaking up bound electrons (represented by curved black arrows to levels 1, 2 and 3). For sufficiently strong probing laser fields, the shake-up electrons can be liberated by tunnelling ionization. The temporal evolution of the tunnelling current will provide information about the inner-atomic electron dynamics that populates and/or depopulates the interrogated shake-up states and the duration of the process that have populated the levels on atto- and femtosecond timescales. Note that the final charge state given in the figure is increased by attosecond tunnelling, while it remains unchanged in the case of attosecond streaking. The observable for streaking is the momentum distribution of liberated electrons, whereas in tunnelling it is the number of ions in different charge states.

Experimental set-up

For a detailed description of the attosecond pump–probe apparatus, see the Supplementary Information. Briefly, the XUV pump originates from high-harmonic generation in a neon gas jet exposed to 300-μJ, 750-nm waveform-controlled laser pulses with a duration of $\tau_L \approx (5.5 \pm 0.5)$ fs, at a repetition rate of 3 kHz. The collinear, linearly polarized XUV and NIR light beams are passed through several filters and reflected by a concentric double-mirror arrangement (Mo/Si multilayer: ~ 9 eV bandwidth at photon energy of 91 eV). The mirror introduces a variable delay between the XUV and the NIR fields, isolates a single or twin XUV pulse (depending on the carrier-envelope phase of the laser pulse) of $\tau_X \approx 250$ as duration[11] by filtering the cut-off part of the harmonic emission spectrum and focuses both beams into a jet of atoms under scrutiny. The absolute delay between the XUV and the NIR signals is determined with an accuracy of better than ±0.5 fs; for details see Supplementary Information. Ions created in the common focus of the two beams are detected by a time-of-flight ion spectrometer (reflectron) that combines high mass resolution ($\Delta m/m \approx 10^{-3}$) with the capability of analysing particles within a micrometre-scale detection volume[17–19]. The target and background pressures were $\sim 10^{-2}$ and $\sim 10^{-8}$ mbar, respectively.

Shake-up and tunnelling

To probe shake-up and light-induced electron tunneling, we ionized neon atoms with our subfemtosecond XUV pulse. Figure 3 shows the level structure and transitions relevant to our experiments. The core shell was not accessed by our XUV photons, and hence Auger decay is absent[20]. The threshold energies for single and double ionization from the outer shell of Ne are 21.56 and 62.53 eV, respectively[21]. The XUV photons produced Ne^{1+} and Ne^{2+} ions with a ratio of

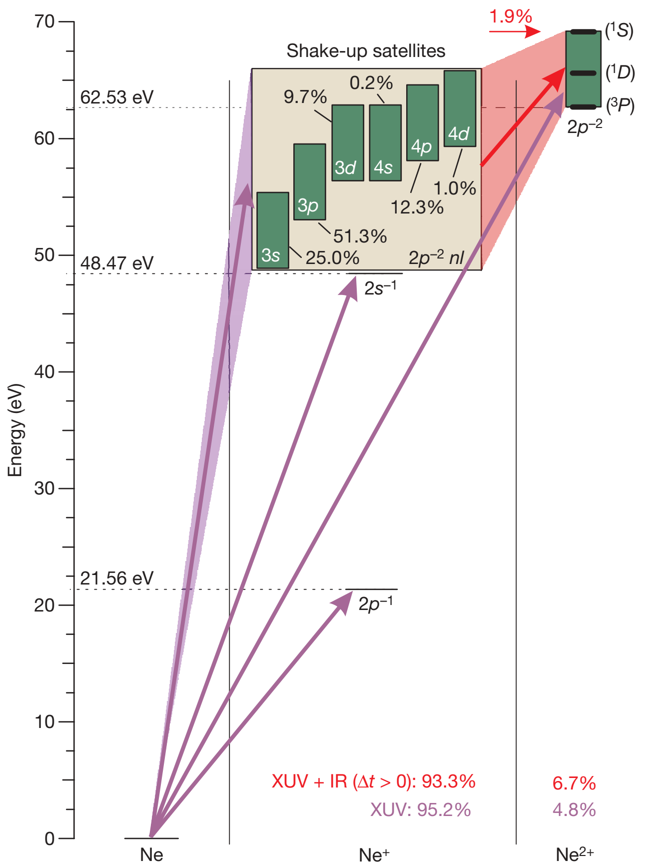


Figure 3 | Energy levels and transitions in Ne^{1+} and Ne^{2+} ions relevant to this study. The plotted levels represent energies required to ionize and possibly excite a neutral atom from its ground state. Absorption of photons of about 91-eV energy can produce singly or doubly charged ions with probabilities of 95.2% and 4.8%, respectively. Owing to shake-up a small fraction of the singly-charged ions is produced in $2p^{-2}nl$ configurations, where the probabilities of different channels are known from electron spectroscopy (see Supplementary Information). Each of these configurations, represented by a green box, consists of $2p^{-2}(^3P)nl$, $2p^{-2}(^1D)nl$, and $2p^{-2}(^1S)nl$ states (the atomic terms in parentheses describe electrons not involved in the interaction). Our few-cycle laser field following the XUV pulse ($\Delta t > 0$) can remove electrons from these shake-up states, thus increasing the probability of double ionization to 6.7%. Depending on the initial $2p^{-2}nl$ state, the doubly ionized ion is left in one of the ^3P, ^1D or ^1S states. On its own, the laser pulse produces only singly charged ions (in configuration $2p^{-1}$) above the detection limit.

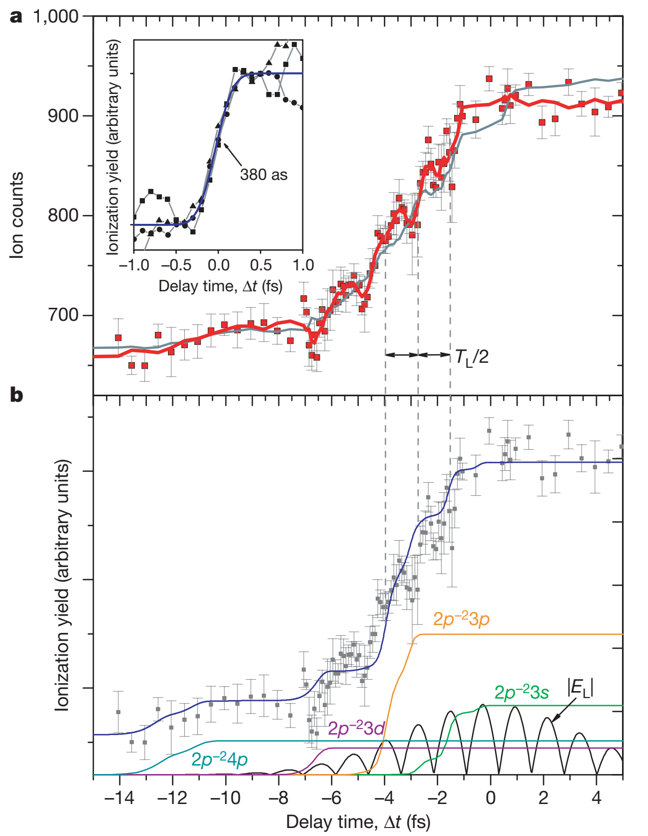


Figure 4 | Ne^{2+} ion yield versus delay between the subfemtoscond XUV pump and the few-cycle NIR probe: experiment and modelling. The peak intensity of the NIR probe was $(7 \pm 1) \times 10^{13}$ W cm^{-2}. **a**, The experimental data were acquired from six delay scans repeated under the same experimental conditions. The signal was accumulated for 3 s at each delay setting. The squares and the error bars show the average and the standard error (s.e.m.) of the results of the six measurements. The thick red line shows the average of five adjacent data points; the thin grey line shows the same as the thick line but recorded with NIR probe pulses of random carrier-envelope phase. In the inset, squares, triangles, and circles depict an ionization step extracted from three different measurements normalized to give the same change in the ionization yield. The solid line shows an error-function fit to the data yielding a rise time of 380 as (full-width at half-maximum, FWHM, of the gaussian function derived from the error function). The zero of delay is set arbitrarily to the centre of the tunnelling process. **b**, Simulation of the pump–probe experiment based on the nonadiabatic theory of tunnel ionization[6]. The thin coloured lines show the calculated fractional ionization yields contributed by electrons liberated from different shake-up states. The thick blue line depicts the overall ionization rate obtained by totalling the fractional rates (and by adding the result to the background measured at large negative delays). The simulations were carried out for a gaussian 250-as XUV pulse and a gaussian 5.5-fs laser pulse with a peak intensity of 7×10^{13} W cm^{-2}. The black solid curve represents the absolute value of the laser field. For discussion of the results, see text.

(19.7 ± 0.5):1 (with ~2,500 Ne^{1+} ions created per second), in good agreement with the results of synchrotron measurements[22]. A few per cent of the Ne^{1+} ions were promoted into $2p^{-2}nl$ (principal quantum number *n*: 3 or 4; quantum orbit *l*: *s*, *p* or *d*) configurations[23] (Fig. 3). These satellite states can only decay radiatively on a picosecond timescale.

Double ionization by the NIR field was not observed at the intensity level chosen. The XUV-generated Ne^{2+} yield was therefore not affected by the NIR probe for $\Delta t \ll -\tau_L$ but was significantly enhanced by the laser field for Δt approaching zero and becoming positive. The NIR-induced Ne^{2+} yield enhancement amounted to (40 ± 4)% of the XUV-produced Ne^{2+} yield at a NIR peak intensity of $(7 \pm 1) \times 10^{13}\,\mathrm{W\,cm^{-2}}$. The absence of this enhancement for $\Delta t \ll -\tau_L$ clearly indicates that the laser sets electrons free from states excited by the XUV pulse as sketched in Fig. 2d. The laser-induced change in the Ne^{2+} yield amounts to some 2% of the XUV-produced Ne^{1+} ions. This implies that a substantial fraction of the population of the $2p^{-2}nl$ shake-up satellites must have been depleted by field ionization.

Figure 4a shows the number of Ne^{2+} ions detected as a function of delay Δt between the XUV pump and NIR probe. Figure 4b compares the prediction of the Yudin–Ivanov theory[6] (lines) with the experimental data (squares). In our modelling the shake-up states were populated instantly during XUV photoionization (for more details, see Supplementary Information). The calculations are in reasonable agreement with our measurements and reveal how the different shake-up states are depleted sequentially by laser-field ionization. The signal starts increasing at large negative delays owing to depletion of the $2p^{-2}4p$ state (relative population ~12%) and the $2p^{-2}3d$ state (~10%) at NIR intensity levels reached some 10 and 6 fs after the peak of the NIR probe ($\Delta t \sim -10$ and -6 fs), respectively. A more dramatic increase in the Ne^{2+} yield is observed as the delay approaches zero, as a consequence of the depletion of the most highly populated $2p^{-2}3p$ (~50%) and $2p^{-2}3s$ (~25%) states. In spite of their relatively high binding energy (~10 and 13 eV, respectively), these states are also depleted before zero delay, that is, by the field oscillation cycles comprised in the trailing edge of the pulse, leaving no room for increasing the Ne^{2+} yield with increasing Δt beyond 0. This main contribution to the Ne^{2+} yield emerges within approximately one and a half wave cycles of the NIR field, $\sim(3/2)T_L = 3\pi/\omega_L$, in several sharp steps that are spaced by $\sim T_L/2$; this clearly shows that field-induced tunnelling is the main cause of the observed increase in the Ne^{2+} yield. This conclusion is also supported by the disappearance of the steps in a pump–probe scan performed with a randomly varying carrier-envelope phase of the NIR probe pulses (grey line in Fig. 4a).

The main shake-up population (residing in the $2p^{-2}3p$ and $2p^{-2}3s$ states) is depleted at NIR intensities corresponding to a Keldysh parameter γ of the order of three. Hence, our experiment verifies not only the existence of light-field-induced tunnelling, as predicted by Keldysh some four decades ago[1], but also confirms the dominant role of this ionization mechanism up to γ values substantially exceeding 1, as predicted recently by Yudin and Ivanov[6]. This conclusion is also backed by numerical solutions of the time-dependent Schrödinger equation (see Supplementary Information).

The steepness of the ionization steps and the dips preceding them in the measured data are not well reproduced by our model, which neglects the influence of electron–electron interactions and that of the strong NIR field on the XUV-induced transitions populating the shake-up states. Recent work[24] and our TDSE simulations (see Supplementary Information) indicate that the influence of the strong laser field on the XUV excitation process may (at least partially) be responsible for this discrepancy.

We feel that these experiments afford profound insight into fundamental electronic processes such as tunnelling and shake-up by contrasting theoretical models with time-domain data. To exploit this potential both (1) accurate models of shake-up in the presence of a strong laser field need to be developed and (2) the temporal resolution needs to be improved by using shorter XUV pulses[25] and improving the signal-to-noise ratio as well as the accuracy of determining the zero of delay. These advances will allow determination of the attosecond temporal evolution of the light-field-induced tunnelling current and they will provide deep insight into the nature of the electron–electron interactions responsible for shake-up.

Once models for shake-up and tunnelling have been tested and verified with attosecond precision, the technique of attosecond tunnelling will provide direct time-domain insight into a wide range of multi-electron dynamics inside atoms and molecules by probing the transient population of excited valence states while these dynamics are unfolding. With improved signal-to-noise and XUV pulse duration, the temporal resolution may potentially approach the atomic unit of time (~24 as). At present, the observed rise time of the Ne^{2+} yield of less than 400 as (which sets a corresponding upper limit on the time it takes the excited electronic states to become populated during XUV ionization and on tunneling; see inset in Fig. 4a) dictates the temporal resolution of our pump–probe approach. In the next section we demonstrate its applicability to probing intra-atomic multi-electron dynamics in real time.

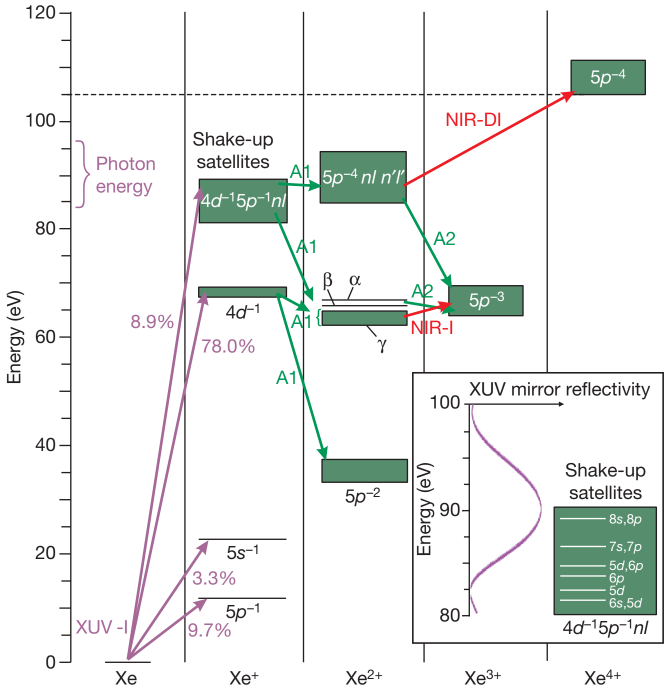


Figure 5 | Energy levels and transitions in xenon ions relevant to the current study. The relative population of states in Xe^{1+} is given for an XUV excitation energy of 90 eV from ref. 30. The XUV light preferably ionizes from the $4d^{-1}$ shell, creating an inner-shell vacancy. A subsequent Auger process (green arrows labelled A1) will follow with 99% probability[26] and is predominantly decaying to Xe^{2+} in configuration $5p^{-2}$. Some of the populated states, α and β (presumably $5s^{-1}5p^{-2}\,7p$ states[28,31]) can further decay to Xe^{3+} via a second Auger process (green arrow labelled A2), leading to triply charged ions. The A1 process also populates states below the threshold for Xe^{3+}. These are denoted by γ and represent $5s^{-1}5p^{-2}\,6p$ as well as $5p^{-3}nl$ configurations. The laser pulse can ionize these states (red arrow labelled NIR-I). Furthermore, a series of $4d^{-1}$ shake-up satellites in Xe^{1+} is also populated—with a small probability—by the XUV pulse. The inset shows the possible configurations[29] together with the XUV mirror reflectivity determining the XUV excitation spectrum. The satellites mainly decay to the α, β, γ states, with a small fraction of the Xe^{2+} population ending up in $5p^{-4}nln'l'$ configurations. These states are short-lived and decay via the A2 process. Before this occurs, the $nln'l'$ electrons can be liberated by the laser field to yield Xe^{4+} in the $5p^{-4}$ state (red arrow labelled NIR-DI).

Multi-electron relaxation

As a first application of the technique of attosecond tunnelling, we captured Auger cascade xenon atoms following excitation by a sub-femtosecond XUV pulse. Figure 5 sketches the relevant energy levels and transitions. Energy-resolved synchrotron measurements have revealed that (1) the 91-eV XUV pulse will preferably liberate electrons from the $4d$ orbital[22], (2) the vacancy decays by subsequent single (A1 in Fig. 5) and cascaded (A1 and A2 in Fig. 5) Auger processes, leading to Xe^{2+} and Xe^{3+}, respectively[26], and (3) the lifetimes of the $4d_{3/2}$ and $4d_{5/2}$ holes are 6.3 ± 0.2 and 5.9 ± 0.2 fs, respectively[27]. These time-integral measurements have hitherto been able to set only a lower limit of 23 fs for the time constant of A2 (ref. 27).

To trace this dynamics in real time, we simultaneously recorded the number of Xe ions emerging in different charged states as a function of Δt. At the laser intensity of $(7 \pm 1) \times 10^{13}\,\mathrm{W\,cm^{-2}}$ used in this experiment, the XUV-induced Xe^{1+} yield is buried in laser-generated background, preventing delay-dependent effects from coming to light in the Xe^{1+} signal. This is not the case for higher charged states. With increasing delay, rapid exponential increase was observed in the Xe^{3+} signal near $\Delta t = 0$ (Fig. 6b), concurrent with a significant decrease of the Xe^{2+} yield. The background in the Xe^{3+} signal arises from the A1–A2 Auger cascade discussed above. From Fig. 5 we infer that the laser-induced increase in the Xe^{3+} yield is due to ionization from the γ-states in Xe^{2+}, which cannot spontaneously decay into Xe^{3+}. This NIR-probe-induced transition (denoted by NIR-I in Fig. 5) yields Xe^{3+} in configuration $5p^{-3}$. Because these are populated by the first Auger process, A1, the exponential increase in the Xe^{3+} signal is the convolution of the A1 decay and the NIR-induced ionization process. No decrease of the enhanced Xe^{3+} yield was observed up to our maximum delay of 300 fs, indicating that the lifetime of the γ states was longer than 1 ps.

Charge-states higher than Xe^{3+} cannot be created with the XUV pulse alone, because the XUV photon energy is below the threshold for Xe^{4+} production (~105 eV). However, with the probe switched on, we did observe Xe^{4+} ions for $\Delta t > -\tau_L$ (Fig. 6a). The Xe^{4+} signal first grows within a few femtoseconds, followed by a longer decay. The Xe^{4+} ions are created by NIR-induced double ionization (denoted by NIR-DI in Fig. 5) from the intermediate doubly excited $5p^{-4}nln'l'$ (Fig. 5) and/or singly excited $5s^{-2}5p^{-1}nl$ (not shown in Fig. 5) states of Xe^{2+}. These states are populated by A1 from the satellites $4d^{-1}5p^{-1}nl$ of the $4d^{-1}$ state upon emission of low-energy electrons[28,29] and emptied by A2 to states of larger binding energy in Xe^{3+}, which cannot be reached by the NIR probe.

From these results we conclude that the exponential rise and decay of the Xe^{4+} signal in Fig. 6a reveal the evolution of the A1 and A2 Auger decays, respectively. The laser-induced double ionization may be either sequential or non-sequential (with the second step induced by recollision of the first electron). In either case, in the first step a binding energy similar to that of the most highly populated shake-up states in the neon experiment must be overcome by tunnel ionization. Hence, the probing (ionization) process may be assumed to be the same as measured in the neon experiment, see Fig. 4a. With this sampling function, fitting the result of a simple rate-equation analysis including the transitions A1 and A2 to the Xe^{4+} data shown in Fig. 6 yields decay times of $\tau_{A1} = 6.0 \pm 0.7$ fs and $\tau_{A2} = 30.8 \pm 1.4$ fs for the A1 and A2 Auger processes, respectively. Both time constants are in accordance with the results of energy-resolved measurements[27].

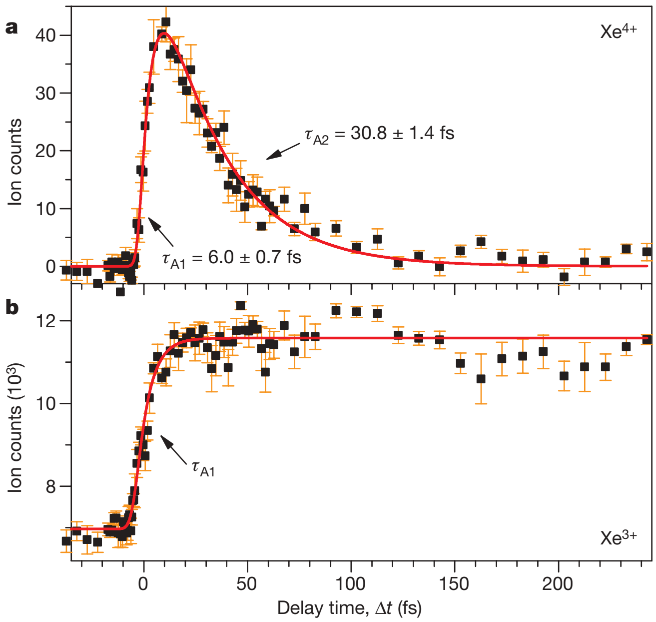


Figure 6 | Xe^{4+} and Xe^{3+} ion yields versus delay between the subfemtosecond XUV pump and the few-cycle NIR probe. The data have been compiled from the results of five delay scans repeated under the same experimental conditions. The signals were accumulated for 20 s at each delay setting. The squares and the error bars show the average and the s.e.m. of the results of these measurements. **a**, Assuming a sampling function identical to the ionization profile measured in the neon experiment (see solid line in Fig. 4a), a double exponential fit (see Supplementary Information) to the measured Xe^{4+} yield versus Δt (solid line) yields the Auger decay times $\tau_{A1} = 6.0 \pm 0.7$ fs and $\tau_{A2} = 30.8 \pm 1.4$ fs. **b**, With the temporal evolution of the A1 process acquired from the Xe^{4+} data and using the fact that the rise of the Xe^{3+} yield with τ comprises both A1 and the laser-induced ionization process, we can obtain the temporal evolution of the laser-induced transition from that of the Xe^{3+} yield. The exponential rise from about 7,000 to 11,500 counts is consistent with a laser-induced ionization time of 5.8 ± 2.5 fs (FWHM). This is comparable to τ_L, indicating that low-order, one-photon and/or two-photon transitions may promote electrons from the γ states of Xe^{2+} to the $5p^{-3}$ states of Xe^{3+} (see Fig. 5). Note that the fluctuations in the Xe^{3+} signal can be largely accounted for by XUV intensity variations, which can be efficiently eliminated by normalization to, for example, the Xe^{2+} ion yield (see Supplementary Information).

Conclusions and outlook

We have reported the observation of light-induced electron tunnelling from atoms in real time. Electrons are found to escape from their atomic binding potential within several subfemtosecond time intervals near the oscillation peaks of the ionizing few-cycle near-infrared laser field. Our results are in good agreement with the predictions of the theory Keldysh put forward four decades ago. The observed subfemtosecond ionization steps provide a powerful means of probing the transient population of short-lived valence electronic states in excited atoms or molecules, offering direct, time-domain access to a wide range of multi-electron dynamics unfolding on an attosecond to femtosecond timescale. Proof-of-principle attosecond tunnelling experiments in neon and xenon demonstrate this potential. Simultaneous implementation of attosecond tunnelling and attosecond streaking spectroscopy along with scaling of the techniques to higher photon energies and shorter X-ray pulse durations will provide unprecedented insight into the transient electronic states of matter.

Received 2 November 2006; accepted 26 January 2007.

1. Keldysh, L. V. Ionization in the field of a strong electromagnetic wave. *Sov. Phys. JETP* **20**, 1307–1314 (1965).
2. Faisal, F. H. M. Multiple absorption of laser photons by atoms. *J. Phys. B* **6**, L89–L92 (1973).
3. Reiss, H. R. Effect of an intense electromagnetic field on a weakly bound system. *Phys. Rev. A* **22**, 1786–1813 (1980).
4. Brabec, T. & Krausz, F. Intense few-cycle laser fields: frontiers of nonlinear optics. *Rev. Mod. Phys.* **72**, 545–591 (2000).
5. Scrinzi, A., Geissler, M. & Brabec, T. Ionization above the coulomb barrier. *Phys. Rev. Lett.* **83**, 706–709 (1999).
6. Yudin, G. L. & Ivanov, M. Yu. Nonadiabatic tunnel ionization: looking inside a laser cycle. *Phys. Rev. A* **64**, 013409 (2001).

7. Baltuska, A. *et al.* Attosecond control of electronic processes by intense light fields. *Nature* **421,** 611–615 (2003).
8. Itatani, J. *et al.* Attosecond streak camera. *Phys. Rev. Lett.* **88,** 173903 (2002).
9. Kitzler, M., Milosevic, N., Scrinzi, A., Krausz, F. & Brabec, T. Quantum theory of attosecond XUV pulse measurement by laser-dressed photoionization. *Phys. Rev. Lett.* **88,** 173904 (2002).
10. Hentschel, M. *et al.* Attosecond metrology. *Nature* **414,** 509–513 (2001).
11. Kienberger, R. *et al.* Atomic transient recorder. *Nature* **427,** 817–821 (2004).
12. Goulielmakis, E. *et al.* Direct measurement of light waves. *Science* **305,** 1267–1269 (2004).
13. Drescher, M. *et al.* Time-resolved atomic inner-shell spectroscopy. *Nature* **419,** 783–787 (2002).
14. Svensson, S., Eriksson, B., Martensson, N., Wendin, G. & Gelius, U. Electron shake-up and correlation satellites and continuum shake-off distributions in x-ray photoelectron spectra of the rare gas atoms. *J. Electron Spectrosc. Related Phenomena* **47,** 327–384 (1988).
15. Aksela, H., Aksela, S. & Kabachnik, N. Resonant and nonresonant Auger recombination. In *VUV and Soft X-Ray Photoionization* (eds Becker, U. & Shirley, D. A.) 401–440 (Plenum, New York, 1996).
16. Istomin, A. Y., Manakov, N. L. & Starace, A. F. Perturbative analysis of the triply differential cross section and circular dichroism in photo-double-ionization of He. *Phys. Rev. A* **69,** 032713 (2004).
17. Schröder, H., Wagner, M., Kaesdorf, S. & Kompa, K. L. Surface-analysis by laser ionization. *Ber. Bunsenges. Phys. Chem.* **97,** 1688–1692 (1993).
18. Wagner, M. & Schröder, H. A novel 4 grid ion reflector for saturation of laser multiphoton ionization yields in a time-of-flight mass-spectrometer. *Int. J. Mass Spectrom.* **128,** 31–45 (1993).
19. Witzel, B., Schröder, H., Kaesdorf, S. & Kompa, K. L. Exact determination of spatially resolved ion concentrations in focused laser beams. *Int. J. Mass Spectrom.* **172,** 229–238 (1998).
20. Larkins, F. P. Charge state dependence of x-ray and Auger electron emission spectra for rare-gas atoms—II. The neon atom. *J. Phys. B* **4,** 14–19 (1971).
21. National Institute of Standards and Technology *Physical Reference Data* ⟨http://physics.nist.gov/PhysRefData/⟩ (1994).
22. Holland, D. M. P., Codling, K., West, J. B. & Marr, G. V. Multiple photoionization in the rare gases from threshold to 280 eV. *J. Phys. B* **12,** 2465–2484 (1979).
23. Becker, U. & Shirley, D. A. Partial Cross Sections and Angular Distributions. In *VUV and Soft X-Ray Photoionization* (eds Becker, U. & Shirley, D. A.) 135–173 (Plenum, New York, 1996).
24. Smirnova, O., Spanner, M. & Ivanov, M. Y. Coulomb and polarization effects in laser-assisted XUV ionization. *J. Phys. B* **39,** 323–339 (2006).
25. Sansone, G. *et al.* Isolated single-cycle attosecond pulses. *Science* **314,** 443–446 (2006).
26. Kämmerling, B., Krässig, B. & Schmidt, V. Direct measurement for the decay probabilities of $4d_j$ hole states in xenon by means of photoelectron-ion coincidences. *J. Phys. B* **25,** 3621–3629 (1992).
27. Penent, F., Palaudoux, J., Lablanquie, P. & Andric, L. Multielectron spectroscopy: the xenon 4d hole double Auger decay. *Phys. Rev. Lett.* **95,** 083002 (2005).
28. Lablanquie, P. *et al.* Photoemission of threshold electrons in the vicinity of the xenon 4d hole: dynamics of Auger decay. *J. Phys. B* **35,** 3265–3295 (2002).
29. Hayaishi, T. *et al.* Manifestation of Kr 3d and Xe 4d conjugate shake-up satellites in threshold-electron spectra. *Phys. Rev. A.* **44,** R2771–R2774 (1991).
30. Becker, U. *et al.* Subshell photoionization of Xe between 40 and 1000 eV. *Phys. Rev. A* **39,** 3902–3911 (1989).
31. Viefhaus, J. *et al.* Auger cascades versus direct double Auger: relaxation processes following photoionization of the Kr 3d and Xe 4d, 3d inner shells. *J. Phys. B* **38,** 3885–3903 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. F. Starace for discussions. We are grateful for financial support from the Volkswagenstiftung (Germany), the Marie Curie Research Training Network XTRA, Laserlab Europe, and a Marie Curie Intra-European Fellowship. F.K. acknowledges support from the FWF (Austria). The research of M.F.K. and M.J.J.V. is part of the research programme of the Stichting voor Fundamenteel Onderzoek der Materie, which is financially supported by the Nederlandse Organisatie voor Wetenschappelijk Onderzoek. This research was supported by the cluster of excellence Munich Centre for Advanced Photonics (www.munich-photonics.de).

Author Contributions M.U., Th.U., M.S. and A.J.V. contributed equally to this work.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.U. (matthias.uiberacker@mpq.mpg.de) or F.K. (ferenc.krausz@mpq.mpg.de).

ARTICLES

Multimodal fast optical interrogation of neural circuitry

Feng Zhang[1], Li-Ping Wang[1], Martin Brauner[2], Jana F. Liewald[2], Kenneth Kay[1], Natalie Watzke[4], Phillip G. Wood[4], Ernst Bamberg[3,4], Georg Nagel[4,5], Alexander Gottschalk[2] & Karl Deisseroth[1]

Our understanding of the cellular implementation of systems-level neural processes like action, thought and emotion has been limited by the availability of tools to interrogate specific classes of neural cells within intact, living brain tissue. Here we identify and develop an archaeal light-driven chloride pump (NpHR) from *Natronomonas pharaonis* for temporally precise optical inhibition of neural activity. NpHR allows either knockout of single action potentials, or sustained blockade of spiking. NpHR is compatible with ChR2, the previous optical excitation technology we have described, in that the two opposing probes operate at similar light powers but with well-separated action spectra. NpHR, like ChR2, functions in mammals without exogenous cofactors, and the two probes can be integrated with calcium imaging in mammalian brain tissue for bidirectional optical modulation and readout of neural activity. Likewise, NpHR and ChR2 can be targeted together to *Caenorhabditis elegans* muscle and cholinergic motor neurons to control locomotion bidirectionally. NpHR and ChR2 form a complete system for multimodal, high-speed, genetically targeted, all-optical interrogation of living neural circuits.

To enable precise perturbation of living circuits, we recently developed an optogenetic[1] technology for genetically targeted, millisecond-timescale optical excitation of neurons[2,3] by employing an algal protein, channelrhodopsin-2 (ChR2)[4], which functions in many cell types[2–9]. However, because ChR2 can only test sufficiency of spike patterns in circuit or behavioural responses and cannot inhibit native spiking, testing the necessity or physiological function of targeted cells remains difficult. The ideal solution would be to complement ChR2 with a hypothetical light-activated hyperpolarizing agent, to permit excitation or inhibition using two light wavelengths while maintaining high temporal precision and genetic targeting. Several state-of-the-art methods[5,10–19] have been developed to kill or inactivate neurons, but none offers the necessary millisecond-precision, rapid reversibility and spectral compatibility for simultaneous use with fast, genetically based, photostimulation techniques such as ChR2.

Finding an optically compatible inhibitor

Our finding that ChR2 could depolarize neurons fast enough to drive precisely timed light-evoked spikes[2] motivated us to search for a complementary high-speed hyperpolarizing agent, to optically silence individual action potentials. ChR2 is highly homologous to ChR1 (ref. 20) and both are homologous to bacteriorhodopsin[4] and the light-driven chloride pumping halorhodopsins[21]. We predicted that the halorhodopsins could be used to optically inhibit neural activity, and we explored halorhodopsins from two strains of archaea, *Halobacterium salinarum* (HsHR) and *Natronomonas pharaonis* (NpHR). We initially characterized the halorhodopsins using *Xenopus laevis* oocytes (see Supplementary Methods for details). Illumination of HsHR- or NpHR-expressing oocytes led to rapid outward currents (Supplementary Fig. 1a). Both HsHR and NpHR have excitation maxima near 580 nm[22,23] (Fig. 1a; the action spectrum of NpHR was measured here in *Xenopus* oocytes using a xenon short arc lamp and narrow-bandwidth 20 nm filters), which, importantly, is red-shifted from the known ChR2 maximum of ~460 nm[4]. This spectral separation indicated to us that ChR2 and a halorhodopsin could be activated independently or in synchrony to effect bidirectional optical modulation of membrane potential.

The dependence of pump current on the extracellular Cl^- concentration for both halorhodopsins was assayed using whole-cell recording. HsHR was found to have a lower extracellular Cl^- affinity than NpHR ($K_{M,NpHR} = 16$ mM, Fig. 1b; $K_{M,HsHR} = 32$ mM) and measured currents showed rapid rundown at low extracellular Cl^- concentrations (Supplementary Fig. 1b), which did not fully recover in darkness. Halorhodopsin pumps are expected to have a very low affinity for Cl^- on the cytoplasmic side, where Cl^- ions are released, because halorhodopsin-mediated chloride pumping can achieve molar concentrations of cytoplasmic Cl^-. We found that the pump currents exhibited more or less linear voltage dependence, and Cl^- current was robust for both halorhodopsins across all physiological voltage regimes (for NpHR see Supplementary Fig. 1c). On the basis of NpHR's higher extracellular Cl^- affinity and stability, we chose to apply NpHR to neurons.

We introduced a mammalian codon-optimized *NpHR* gene fused with enhanced yellow fluorescent protein (*NpHR–EYFP*) into cultured rat hippocampal CA3/CA1 neurons using lentiviruses carrying the ubiquitous *EF1α* promoter (*EF1α-NpHR–EYFP*). Cells expressing NpHR–EYFP maintained robust expression (Fig. 1c) for weeks after infection. In voltage-clamp experiments, illumination of NpHR–EYFP cells with yellow light (bandwidth 573–613 nm using a Semrock filter FF01-593/40-25 and 300 W xenon lamp) induced rapid outward currents (Fig. 1d, upper panel) with a peak level of 43.8 ± 25.9 pA and a steady-state level of 36.4 ± 24.4 pA (mean ± s.d. reported throughout this paper, unless otherwise stated, $n = 15$; Fig. 1e). The relatively small difference between the peak and steady-state currents is indicative of rare deprotonation of the NpHR Schiff base during the pump cycle[22]. The rise time from light onset to 50% of the peak current was consistent across cells

[1]Department of Bioengineering, Stanford University, Stanford, California 94305, USA. [2]Institute of Biochemistry, and [3]Institute of Biophysical Chemistry, Department of Biochemistry, Chemistry and Pharmacy, Johann Wolfgang Goethe-University, Frankfurt Biocenter N220, Max-von-Laue Straβe 9, D-60438 Frankfurt, Germany. [4]Max-Planck-Institute of Biophysics, Max-von-Laue-Straβe 3, D-60438 Frankfurt am Main, Germany. [5]University Wuerzburg, Botanik I, Julius-von-Sachs-Platz 2, D-97082 Wuerzburg, Germany.

(6.0 ± 1.0 ms; Fig. 1f) with rise and decay time constants of τ_{on} = 6.1 ± 2.1 ms and τ_{off} = 6.9 ± 2.2 ms, respectively. Light-evoked responses were never seen in cells expressing EYFP alone (data not shown). In current-clamp experiments, NpHR–EYFP neurons exhibited light-evoked hyperpolarization (Fig. 1d; lower panel) with an average peak of 14.7 ± 6.9 mV and a steady-state of 12.1 ± 6.6 mV (Fig. 1g). The delay from light onset to 50% of hyperpolarization peak was 26.0 ± 8.6 ms (Fig. 1f), and the rise and decay time constants were τ_{on} = 35.6 ± 15.1 ms and τ_{off} = 40.5 ± 25.3 ms, respectively. To test whether NpHR-mediated hyperpolarization could inhibit neuronal firing, we injected current-clamped neurons with a 200 pA current step for 2 s to evoke robust spike firing; concurrent light delivery abolished the evoked activity (Fig. 1h). Therefore, NpHR can indeed serve as an optical inhibitor of neuronal activity.

We next verified that NpHR functions in an unperturbed intracellular environment. We co-expressed NpHR–EYFP and ChR2 tagged with a red fluorescent protein variant (ChR2–mCherry) in cultured hippocampal neurons (Fig. 2a), and we probed NpHR function using cell-attached recordings with ChR2 photostimulation to drive reliable spike trains. Indeed, whereas trains of blue light pulses were able to evoke action potentials through the activation of ChR2, concomitant yellow light illumination abolished spike firing in both cell-attached and subsequent whole-cell recoding modes (Fig. 2b). After achieving the whole-cell configuration, voltage-clamp recording showed that independent exposure to yellow or blue light led to outward or inward photocurrents, respectively (Fig. 2c). This further confirms that ChR2 and NpHR can be combined to achieve bidirectional, independently addressable modulation of membrane potential in the same neuron. In addition, we found that NpHR inhibitory function does not require a specific pipette chloride concentration under these recording conditions, because NpHR-mediated inhibition is robust across a range of relevant whole-cell pipette chloride concentrations (4–25 mM) and physiologically negative resting potentials (Supplementary Fig. 2)—as expected from the fact that NpHR has evolved to deliver chloride ions to molar levels in the archaeal intracellular milieu.

Effects of NpHR on neuronal physiology

We conducted extensive controls to test whether heterologous expression of NpHR in neurons would alter the membrane properties or survival of neurons. Lentiviral expression of NpHR for at least 2 weeks did not alter neuronal resting potential (−53.1 ± 6.3 mV for $NpHR^+$ cells, −57.0 ± 4.8 mV for $NpHR^-$ cells, and −56.7 ± 5.7 mV for $NpHR^+$ cells exposed to yellow light for 10 min followed by a delay period of 1 day; Supplementary Fig. 2a, n = 12 each) or membrane resistance (114.5 ± 34.1 MΩ for $NpHR^+$ cells, 108.9 ± 20.1 MΩ for $NpHR^-$ cells, and 111.4 ± 23 MΩ for the light-exposed $NpHR^+$ cells; Supplementary Fig. 2b, n = 12 each). These electrical measurements indicate that NpHR had little basal electrical activity or passive current-shunting ability and did not compromise cell health. We also tested the dynamic electrical properties of neurons with and without NpHR. There was no significant difference in the number of spikes evoked by 500 ms of a 300-pA current injection (7.5 ± 2.8 for $NpHR^+$ neurons, 10.7 ± 7.9 for $NpHR^-$ neurons, and 9.3 ± 5.1 for the light-exposed $NpHR^+$ neurons; Supplementary Fig. 2c). Finally, we stained live $NpHR^+$ neurons (with and without light exposure) and $NpHR^-$ neurons with the membrane-impermeant DNA-binding dye propidium iodide to assess cell survival. NpHR expression did not affect the percentage of neurons that took up propidium iodide

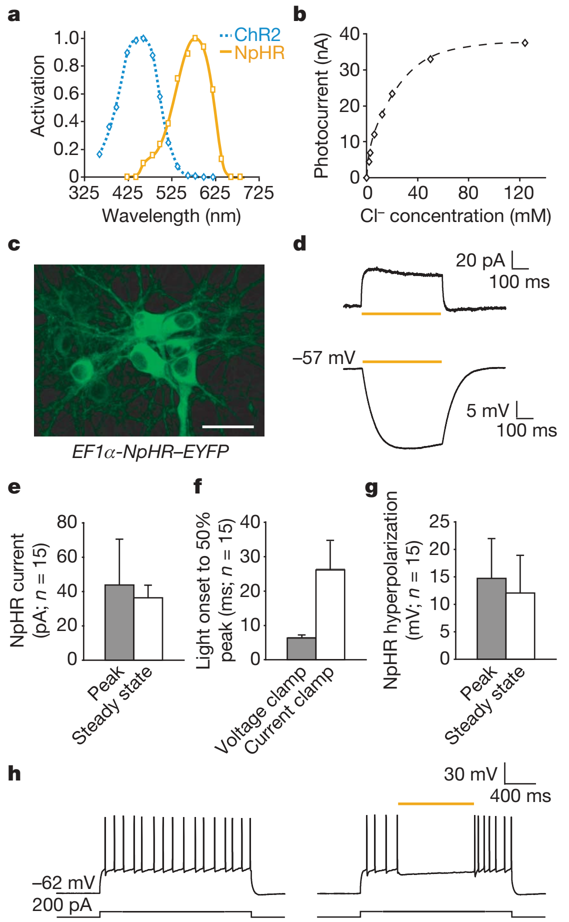


Figure 1 | Electrophysiological properties of NpHR in oocytes and hippocampal neurons. a, Action spectrum of NpHR in oocytes held at −50 mV, determined by 20 nm narrow bandwidth interference filters. The ChR2 action spectrum is provided for comparison[4]. **b**, Extracellular [Cl^-] dependence of NpHR pump currents indicates a simple Michaelis–Menten type saturation with a K_M of 16 mM (dashed fit curve). **c**, Hippocampal neurons expressing NpHR–EYFP (scale bar 50 μm). **d**, Yellow light (593 nm)-induced outward photocurrent in neurons (top panel, voltage clamp) and membrane hyperpolarization (bottom panel, current clamp); illumination duration is indicated by the yellow bar. **e**, NpHR peak versus steady-state current (mean ± s.d.; n = 15). **f**, Latency of NpHR activity measured from light onset to 50% of the peak current or hyperpolarization (mean ± s.d.; n = 15). **g**, NpHR peak versus steady-state membrane hyperpolarization (mean ± s.d.; n = 15). **h**, Illumination with yellow light potently inhibited neuronal firing.

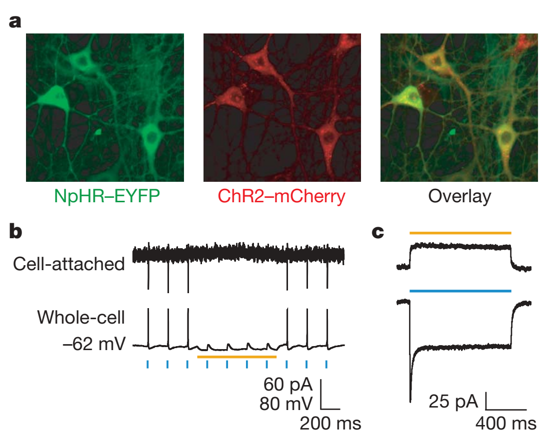


Figure 2 | Combining NpHR with ChR2 for noninvasive optical control. a, Hippocampal neurons co-expressing NpHR–EYFP under control of the *EF1α* promoter and ChR2–mCherry under control of the *synapsin I* promoter. **b**, Cell-attached and whole-cell recording of neurons co-expressing NpHR–EYFP and ChR2–mCherry. Action potentials are evoked by brief pulses of blue light (473 nm, 15 ms per pulse; length of blue bars is not to scale for ease of visualization). Simultaneous illumination with yellow light inhibited spike firing. **c**, Voltage-clamp recording from a single neuron co-expressing NpHR–EYFP and ChR2–mCherry, showing independently addressable outward and inward photocurrents in response to yellow and blue light, respectively.

(13/240 for NpHR$^+$ cells, 7/141 for NpHR$^-$ cells, and 10/205 for the light-exposed NpHR$^+$ cells; Supplementary Fig. 2d, $P > 0.999$ by χ^2 test). These experiments indicate that NpHR expression does not significantly affect the health or basal electrical properties of neurons.

NpHR operates over a range of timescales

To quantify the capabilities of NpHR for inhibiting neural activity, we first determined the tunability of NpHR efficacy with different intensities of delivered light. Using a 200 pA current step, which drove reliable action potential trains, we found that maximal light intensity of 21.7 mW mm^{-2} under a ×40, 0.8 numerical aperture water-immersion objective inhibited 98.2 ± 3.7% of the spikes (Fig. 3a and b). Using 33% or 50% of the full light intensity inhibited 74.9 ± 22.2% and 87.3 ± 13.5% of spikes, respectively (Fig. 3b). With steady current injection, lower intensities of light were effective for a shorter period of time; the delays from light onset to the first escaped spike under 33%, 50% and 100% light intensity were 533.65 ± 388.2 ms, 757.5 ± 235.5 ms and 990.5 ± 19.1 ms, respectively (Fig. 3c). Therefore, inhibition will be more effective early in the light pulse, presumably owing to the slight inactivation of NpHR. Except where otherwise noted, we conducted the rest of the experiments with 21.7 mW mm^{-2} yellow light delivered to the neurons.

Using trains of brief current pulses to generate spike trains, we tested whether NpHR could mediate both long-term inhibition (to emulate lesions on the timescale of seconds to minutes) and short-term inhibition (to modify spike firing on the millisecond timescale). We first tested long-term inhibition over 10 min by injecting 300-pA current pulses at 5 Hz to drive steady action potential firing. Concurrent yellow light was delivered continuously for 10 min. NpHR-mediated inhibition of spike trains was highly stable, remaining effective over many minutes (Fig. 3d). Almost all (99.0 ± 1.9%) spikes were inhibited within the first two minutes, and over 90% of spikes were inhibited for up to 8 min ($n = 5$; Fig. 3d, inset). The slight decrease in efficacy is probably due to accumulation of non-functional NpHRs with a deprotonated Schiff base over long periods of light exposure[24]. Although natural reprotonation of the Schiff base is slow, any non-functional NpHRs can be readily and quickly restored by brief illumination with blue light[24] (Supplementary Fig. 2h).

NpHR inhibits with single-spike precision

We next tested whether NpHR activation might allow the 'knockout' of single action potentials. Because the fast photocurrent of ChR2 enables brief pulses of blue light to drive reliable action potential trains[2], we used brief pulses of yellow light to test NpHR-mediated

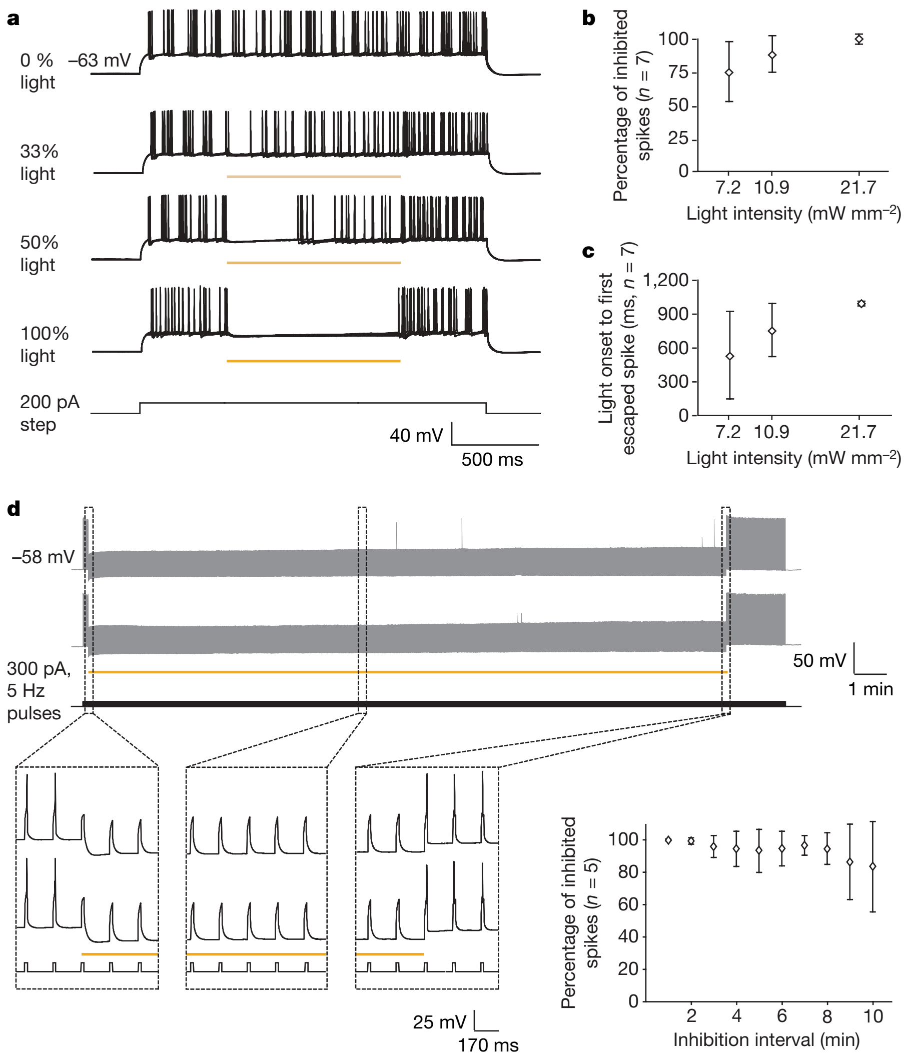


Figure 3 | NpHR mediates tunable neuronal inhibition over a range of timescales. a, Increasing intensities of light lead to stronger inhibition of spikes evoked by 200 pA of current injection (21.7 mW mm^{-2} at 100% intensity). **b,** Per cent of spikes inhibited at three different light intensities (1 s of light exposure, mean ± s.d.; $n = 7$). **c,** Duration of effective inhibition as measured by the delay from light onset to the first escaped spike (mean ± s.d.; $n = 7$). **d,** Ten min continuous illumination (two successive sweeps) robustly inhibited neuronal spike trains elicited by current pulse injection (300 pA, 5 Hz). Dashed boxes show the inhibition on an expanded timescale. Per cent efficacy of inhibition over 10 min measured by breaking down the 10 min interval into 1 min bins (inset, mean ± s.d.; $n = 5$).

inhibition. To assess the temporal resolution of spike inhibition, we attempted to inhibit pairs of spikes in action potential trains of 5, 10, 20 and 30 Hz (Fig. 4a) and found that single spikes could be reliably inhibited even from within spike trains (Fig. 4a). To define the temporal precision of NpHR we inhibited selected pairs of spikes with a range of inter-spike temporal delays. Both closely timed and temporally separated spike pairs could be reliably inhibited, while sparing spikes between the targeted times ($n = 6$, Fig. 4a–c). Over spike rates of 5 to 30 Hz, the closely timed spikes could be selectively inhibited with a probability of 0.95 or greater. We also found that by giving trains of millisecond-scale yellow light pulses, it was straightforward to simulate barrages of inhibitory postsynaptic potential (IPSP)-like events with precise, reliable timing and amplitudes, from 5 to 100 Hz (Fig. 4d). Because NpHR is a Cl^- pump and not a channel, the light-driven inhibition acts by shifting the membrane potential and will not contribute to shunting or input resistance changes. Indeed, we found that whereas the $GABA_A$ chloride channel agonist muscimol significantly decreased neuronal input resistance, NpHR activation had no detectable effect on the input resistance (Fig. 4e, f).

Because both ChR2 and NpHR can be activated with high temporal precision using millisecond-scale blue or yellow light pulses, we next asked if it would be possible to drive both proteins in intermingled, temporally precise patterns to non-invasively activate or inhibit single identified action potentials with light in the same experiment, or even in the same cell. Indeed, cell-attached and whole-cell recordings in hippocampal pyramidal neurons revealed that precisely patterned trains of yellow and blue light pulses can be used to evoke and inhibit neural activity with single-spike precision (Fig. 5a), and that NpHR can be used to override multiple preselected ChR2-driven spikes at identified positions in prolonged spike trains (Fig. 5a).

Application of NpHR in the mammalian brain

A crucial feature of ChR2 is that it can be functionally expressed in the mammalian brain without exogenous delivery of its required cofactor all-*trans*-retinal (ATR)[3], presumably owing to the presence of endogenous retinoids in the mammalian brain. We anticipated that this unique and important property could extend to NpHR, which also employs an ATR chromophore. To test this possibility, we delivered lentiviruses carrying NpHR–EYFP under the neuronal *CaMKIIα* promoter into the hippocampus of the adult mouse. Neurons throughout the hippocampus exhibited stable expression of NpHR–EYFP, as indicated by robust EYFP fluorescence (CA3 subfield shown; Fig. 5b, left

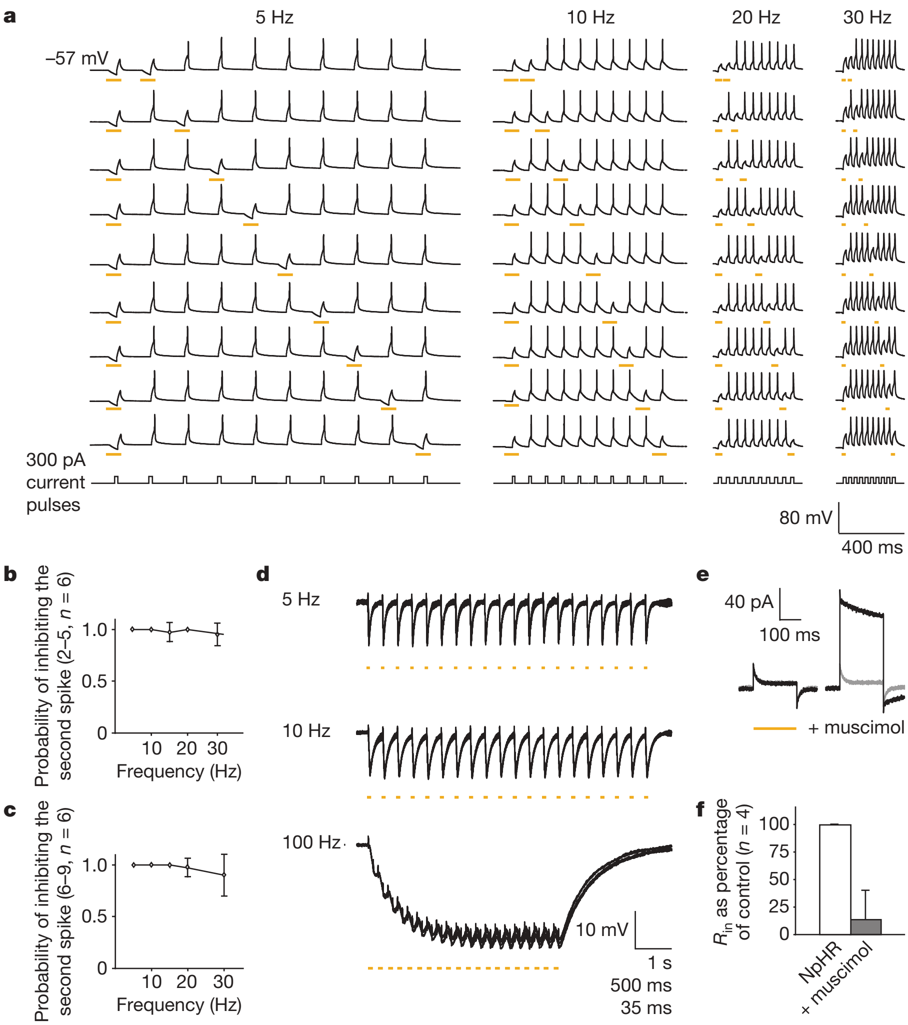


Figure 4 | Temporal precision of NpHR-mediated inhibition. a, Inhibition of spike pairs at different frequencies. Trains of ten spikes were evoked by 300 pA current injection pulses, and pairs of light pulses with variable temporal separation were used to inhibit the selected spikes. **b,** Probability of inhibiting the second spike in the pair, as a function of spike frequency, for short separations (spikes 2–5 in the spike train, mean ± s.d.; $n = 6$). **c,** Probability of inhibiting the second spike in the pair, as a function of spike frequency, for longer separations (spikes 6–9 in the spike train, mean ± s.d.; $n = 6$). **d,** NpHR can be employed to mimic precisely timed IPSP barrages at different frequencies; five successively acquired traces are overlaid for each frequency. **e,** Voltage-clamp recording of NpHR–EYFP-expressing neurons. Test pulses (5 mV, 200 ms) were used to test for changes in input resistance (R_{in}) during yellow light illumination or 50 nM muscimol application. Grey traces show the control condition. **f,** Summary plot showing that NpHR activation did not affect R_{in} (100.2 ± 1.15% of control, mean ± s.d.; $n = 4$), whereas muscimol application significantly decreased R_{in} (14 ± 26.8% of control, mean ± s.d.; $n = 4$).

panel). NpHR–EYFP cells in acute hippocampal slices exhibited voltage-clamp photocurrents similar to those observed in cultured neurons (data not shown), and current clamp recording of NpHR–EYFP neurons revealed that temporally precise patterns of spike inhibition could be achieved readily, as in dissociated culture (Fig. 5b, right panel). No exogenous cofactors were delivered at any point, indicating that NpHR can be functionally applied to mammalian systems *in vivo*.

Combining NpHR and ChR2 with imaging and behaviour

We next combined the NpHR/ChR2 system, expressed in living mammalian neural circuitry, with fura-2 calcium imaging in an all-optical experiment. Lentiviruses carrying ChR2–mCherry under the neuron-specific *CaMKIIα* promoter and NpHR–EYFP under the *EF1α* promoter were injected into the brain of postnatal day 4 (P4)

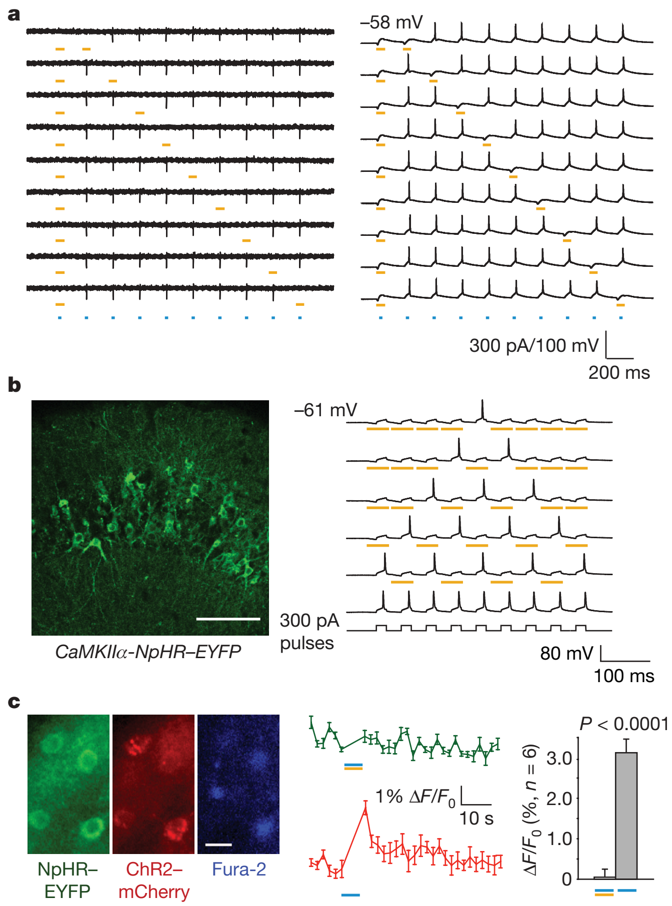


Figure 5 | Bidirectional optical neural control and *in vivo* implementation. **a**, Cell-attached (left panel) and whole-cell (right panel) recording of cultured hippocampal neurons co-expressing ChR–mCherry and NpHR–EYFP. Action potentials were evoked by trains of blue light pulses (5 Hz trains, 15 ms pulse width, blue bars). NpHR-mediated inhibition was co-administered by brief yellow light pulses (50 ms pulse width, yellow bars). **b**, Confocal image from an acute mouse brain slice showing membrane-localized NpHR–EYFP expression in the hippocampal CA3 subfield (left panel; scale bar, 150 μm). Current-clamp recording at right shows NpHR-mediated inhibition of specific spikes during a train of action potentials evoked by pulsed current injection (300 pA, 20 Hz). **c**, Left panel, epifluorescence images of cortical neurons triple-labelled with NpHR–EYFP, ChR2–mCherry and Fura-2. Scale bar, 20 μm. Middle panel, simultaneous illumination of cells co-expressing NpHR–EYFP and ChR2–mCherry with steady yellow (continuous illumination, 6 s) and pulsed blue light (50 pulses at 15 ms per flash, 10 Hz) prevented $[Ca^{2+}]_i$ transients (top trace, mean ± s.e.; $n = 6$). Subsequent photostimulation of the same cells with blue light pulses (50 pulses at 15 ms per flash, 10 Hz) evoked reliable $[Ca^{2+}]_i$ transients (bottom trace). Right panel, summary graph of photostimulation-induced fluorescence changes (mean ± s.e.; $n = 6$ ChR2-activated triple-labelled cells).

mouse pups; acute cortical slices were prepared at P10–P14 and labelled with fura-2-acetoxymethyl. In neurons co-expressing ChR2–mCherry and NpHR–EYFP (Fig. 5c, left panel), initial simultaneous illumination with both blue and yellow light did not lead to intracellular calcium concentration ($[Ca^{2+}]_i$) transients, whereas subsequent pulsed blue light alone in the same neurons evoked ChR2-triggered $[Ca^{2+}]_i$ transients (Fig. 5c, middle panel), demonstrating that NpHR and ChR2 can be integrated to achieve multimodal, bidirectional control of neural activity in intact tissue. In the same imaged cells where ChR2 stimulation led to a 3.1 ± 0.3% increase in fluorescence intensity ratios $\Delta F/F_0$, the combination of NpHR and ChR2 activation resulted in a 0.0 ± 0.2% effect on $\Delta F/F$ ($n = 6$, $P < 0.0001$; Fig. 5c, right panel). Yellow illumination alone had no detectable effect on $[Ca^{2+}]_i$ (not shown). Because not all targeted cells are necessarily affected to the same degree (Fig. 1), this optical system could complement electrophysiology to probe successful modulation of the targeted cell population.

Combining ChR2 and NpHR with calcium imaging provides an all-optical system for interrogation of neural circuits. We next asked if this system could robustly control animal behaviour *in vivo*. We first expressed an NpHR–ECFP fusion protein in the body wall muscles of the nematode *Caenorhabditis elegans* using the muscle-specific myosin promoter (*Pmyo-3*). ECFP fluorescence could be readily observed throughout muscle cells and membranous muscle arm extensions (Fig. 6a and Supplementary Fig. 3a). Because nematode worms (unlike mammals) seem not to have sufficient levels of endogenous retinoids[6], transgenic animals expressing NpHR in muscle were grown in medium containing ATR. Whole-cell voltage-clamp recordings from dissected muscles indeed demonstrated light-evoked outward currents (265 ± 82 pA, $n = 9$; Fig. 6a). To test effects on muscle activity, swimming behaviour in liquid medium was analysed. Consistent with the photocurrents observed, photoactivation of NpHR immediately (within ~150 ms) and essentially completely arrested swimming behaviour (Fig. 6a, b and Supplementary Movie 1). As controls, we used transgenic animals raised in the absence of ATR, and wild-type animals raised with and without ATR. Importantly, robust paralysing effects of light were observed, but consistently only in transgenic animals raised in the presence of ATR. When muscle-expressing animals were illuminated for 1 s, they quickly returned to their natural swimming rate after light-stimulus termination (Fig. 6b). When NpHR was activated in muscle for 10 s (Fig. 6b), animals remained uncoordinated for prolonged periods (up to 40 s), before a full recovery became apparent and normal swimming commenced (Supplementary Fig. 3c).

We next targeted NpHR to a specific class of genetically defined neurons *in vivo*. We expressed NpHR–ECFP in cholinergic motor neurons using the vesicular acetylcholine transporter promoter (*Punc-17*)[25]. When illuminated for 1 or 10 s, respectively, these animals also strongly reduced or essentially stopped swimming behaviour (Fig. 6b and Supplementary Movie 2). These animals, in contrast to the muscle-targeted individuals, recovered to normal swimming behaviour immediately. This may indicate that neurons have more powerful Cl^- homeostasis mechanisms than muscles, though in all cases full recovery was observed, as expected from the lack of toxicity observed in mammalian neurons (Supplementary Fig. 2).

When illuminated on solid agar substrate, transgenic animals expressing NpHR either in muscle or in cholinergic motor neurons, exhibited rapid inhibition of movement and relaxed their bodies, resulting in overall elongation by up to 9% within ~600 ms of illumination (Fig. 6c and Supplementary Movies 3, 4 and 5). Finally, we found that ChR2 and NpHR could be driven simultaneously in *C. elegans* as well. With either muscle or targeted cholinergic neuron expression (using the *Pmyo-3* or *Punc-17* promoters, respectively), NpHR rapidly and reversibly counteracted the shortening behaviour observed with ChR2 alone (Fig. 6d and Supplementary Movies 6 and 7). These experiments demonstrate for the first time that acetylcholine release can be efficiently triggered from *C. elegans* motor neurons using

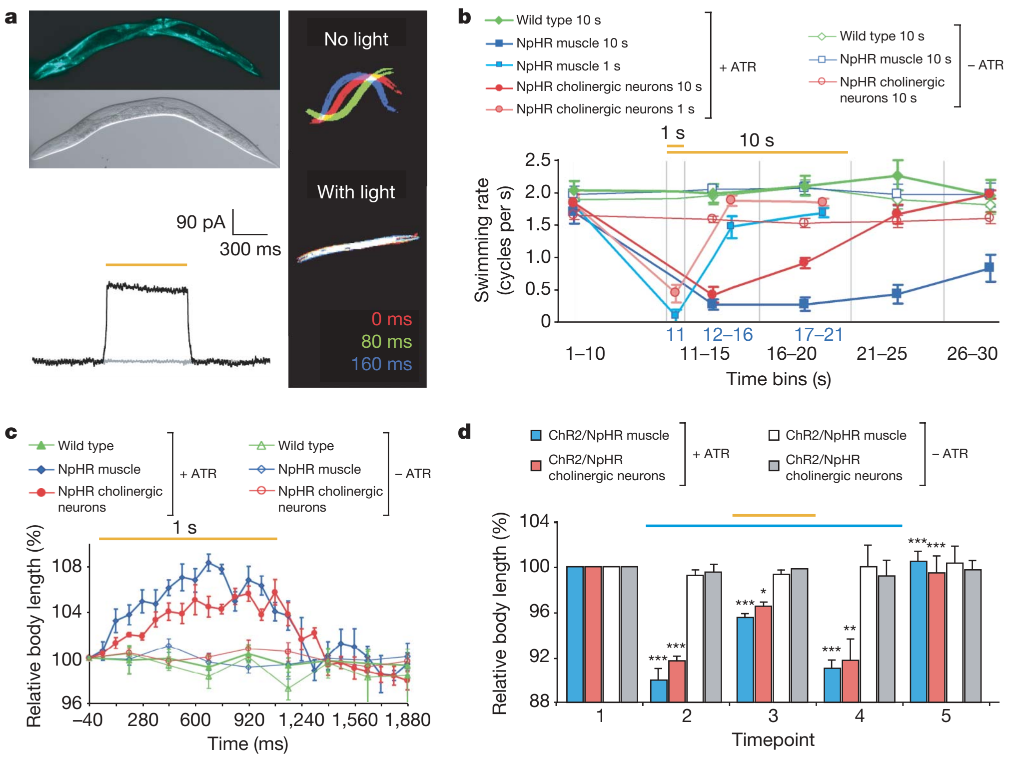


Figure 6 | Bidirectional optical control of *C. elegans*. **a**, Epifluorescence image showing *Pmyo-3*-mediated (transgene *zxEx30*) NpHR–ECFP expression in the body wall muscles of *C. elegans* (top left panels; see Supplementary Fig. 3a for higher magnification). Muscle voltage-clamp trace (bottom left panel) showing photocurrent in transgenic *C. elegans* expressing NpHR–ECFP (transgene *zxEx30*) and raised in the presence of ATR (black trace); note lack of response in transgenic animal raised in the absence of ATR (grey trace). NpHR activation targeted to muscle using *Pmyo-3* blocked swimming behaviour in liquid medium (right panels, transgene *zxEx29*). Animal postures from three consecutive movie frames (frame rate 12.5 Hz), either with or without NpHR photoactivation, were superimposed. **b**, Effect of 10 s illumination on swimming rate (mean ± s.e.; $n = 10$ for each set) in wild-type controls (green), animals expressing NpHR in muscles (transgene *zxEx29*; blue), or cholinergic motor neurons (transgene *zxEx33*; red). The number of swimming cycles per second was counted in bins of 10 s or 5 s intervals; illumination is shown as the orange bar. A briefer 1 s illumination protocol was also used, revealing rapid inhibition during illumination and faster recovery by comparison with the 10 s illumination (see also Supplementary Fig. 3c for full muscle recovery analysis). **c**, Effect of 1 s illumination on body length (mean ± s.e.; $n = 5$–6 for each set). Movies were taken from transgenic worms expressing NpHR in muscles (transgene *zxEx29*) or cholinergic motor neurons (transgene *zxEx33*). Plot shows the body length during the first frame before illumination, the 13 frames during the illumination, and the next 11 frames after illumination ended. Colour-coding and controls were as in **b**. **d**, Combined ChR2/NpHR expression in muscle cells (transgene *zxEx32*) or cholinergic motor neurons (transgene *zxEx34*) of worms. NpHR activation significantly reversed the muscle contraction caused by ChR2 activation (mean ± s.e.; $n = 6$ per condition; $*P < 0.05$; $**P < 0.01$; $***P < 0.001$; between consecutive time points 1 s apart; *t*-test), but not in animals raised in the absence of ATR.

ChR2, and that ChR2 and NpHR work well together in nematodes as well as mammals.

Discussion

The NpHR/ChR2 system enables rapid bidirectional control of neurons on the timescale of milliseconds, thus enabling emulation or alteration of the neural code. These fast genetically based, neural-spike-controlling technologies powerfully augment existing tools for interrogating neural systems[26–33]. Although other optical neural modulators typically require the addition of exogenous chemicals[10,18,34,35], both NpHR and ChR2 can be functionally expressed and operate at high speed in the mammalian brain without necessitating cofactor addition. Moreover, NpHR and ChR2 function in *C. elegans* as well (Fig. 6) after simple dietary ATR supplementation[9]. When combined with imaging[36] or other measures in intact tissue or freely moving animals, the NpHR/ChR2 system provides the capability to directly and causally link precisely defined patterns of neural activity with specific circuit behaviours.

The ability to use light to inhibit[5,10,37] or activate[2,34,35,38–42] neurons may have practical applications beyond basic science investigation, as the NpHR/ChR2 system could be genetically targeted to specific classes of neurons or other excitable cells involved in disease processes to enable highly precise optical therapeutic treatments. For example, in Parkinson's disease electrode-based deep brain stimulation can be therapeutic for symptomatic relief but also gives rise to side effects. Delivery of these optogenetic tools targeted by cell-type-specific promoters to distinct disease-related neuronal types may provide a more precise alternative with fewer side effects and will lead to physiological insights into animal models of disease.

Whether applied in basic science or clinical applications, the spectral separation between the NpHR and ChR2 activation maxima permits both sufficiency and necessity testing in elucidation of the roles of specific cell types in high-speed intact circuit function. Indeed, integration of GFP-based probes and fura-2 with the NpHR/ChR2 neural control system delivers a powerful and complementary triad of technologies to identify, observe and control intact living neural circuitry with light.

METHODS SUMMARY

DNA constructs. Plasmid constructs were prepared according to standard molecular biology protocols. The lentiviral backbone was provided as a gift by P. Osten[43]. A list of *C. elegans* transgenes used can be found in Supplementary Table 1.

Lentivirus preparation and stereotactic injection. Lentiviruses were prepared as previously described[2] (a step-by-step instruction can be found in the methods). For electrophysiology, 1 µl of virus solution was injected into the hippocampal CA3 region of 5–7-week-old C57BL/6 mice (anteroposterior, −2.0 mm from bregma; lateral, 2 mm; ventral, 2.2 mm). For calcium imaging younger animals were used (postnatal day 4 or 5).

Electrophysiology and calcium imaging. Recordings in oocytes[4,20] and neurons[2] were carried out as previously described. Detailed composition of intra- and

extracellular solutions can be found in the methods. For calcium imaging, Fura-2-AM was loaded into the regions of interest as previously described[44]. Excitation of NpHR and ChR2 were accomplished using 593 nm and 437 nm excitation filters, respectively. The 300 W Lambda DG-4 (Sutter Instruments) was used for illumination and rapid wavelength switching. Frames were collected at 0.5 Hz. $[Ca^{2+}]_i$ changes were detected in single cells and expressed as the fluorescence intensity ratios $\Delta F/F_0$ with the resting fluorescence value F_0 determined at the beginning of each experiment. Patch clamping of *C. elegans* muscle was performed as previously described[6], at a holding potential of -40 mV.

***C. elegans* stimulation and electrophysiology.** Transgenic worms were generated using standard transgenic techniques. *C. elegans* were grown in the presence of ATR, as described previously[6] except that the ATR concentration was reduced to 5 μM. NpHR activation was achieved using a 50 W mercury lamp and band-pass filters. For co-activation of ChR2, a 473 nm laser was used.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 23 December 2006; accepted 14 March 2007.

1. Deisseroth, K. *et al.* Next-generation optical technologies for illuminating genetically targeted brain circuits. *J. Neurosci.* **26**, 10380–10386 (2006).
2. Boyden, E. S., Zhang, F., Bamberg, E., Nagel, G. & Deisseroth, K. Millisecond-timescale, genetically targeted optical control of neural activity. *Nature Neurosci.* **8**, 1263–1268 (2005).
3. Zhang, F., Wang, L. P., Boyden, E. S. & Deisseroth, K. Channelrhodopsin-2 and optical control of excitable cells. *Nature Methods* **3**, 785–792 (2006).
4. Nagel, G. *et al.* Channelrhodopsin-2, a directly light-gated cation-selective membrane channel. *Proc. Natl Acad. Sci. USA* **100**, 13940–13945 (2003).
5. Li, X. *et al.* Fast noninvasive activation and inhibition of neural and network activity by vertebrate rhodopsin and green algae channelrhodopsin. *Proc. Natl Acad. Sci. USA* **102**, 17816–17821 (2005).
6. Nagel, G. *et al.* Light activation of channelrhodopsin-2 in excitable cells of *Caenorhabditis elegans* triggers rapid behavioral responses. *Curr. Biol.* **15**, 2279–2284 (2005).
7. Ishizuka, T., Kakuda, M., Araki, R. & Yawo, H. Kinetic evaluation of photosensitivity in genetically engineered neurons expressing green algae light-gated channels. *Neurosci. Res.* **54**, 85–94 (2006).
8. Bi, A. *et al.* Ectopic expression of a microbial-type rhodopsin restores visual responses in mice with photoreceptor degeneration. *Neuron* **50**, 23–33 (2006).
9. Schroll, C. *et al.* Light-induced activation of distinct modulatory neurons triggers appetitive or aversive learning in *Drosophila* larvae. *Curr. Biol.* **16**, 1741–1747 (2006).
10. Banghart, M., Borges, K., Isacoff, E., Trauner, D. & Kramer, R. H. Light-activated ion channels for remote control of neuronal firing. *Nature Neurosci.* **7**, 1381–1386 (2004).
11. Ehrengruber, M. U. *et al.* Activation of heteromeric G protein-gated inward rectifier K^+ channels overexpressed by adenovirus gene transfer inhibits the excitability of hippocampal neurons. *Proc. Natl Acad. Sci. USA* **94**, 7070–7075 (1997).
12. Ibanez-Tallon, I. *et al.* Tethering naturally occurring peptide toxins for cell-autonomous modulation of ion channels and receptors *in vivo*. *Neuron* **43**, 305–311 (2004).
13. Isles, A. R. *et al.* Conditional ablation of neurones in transgenic mice. *J. Neurobiol.* **47**, 183–193 (2001).
14. Johns, D. C., Marx, R., Mains, R. E., O'Rourke, B. & Marban, E. Inducible genetic suppression of neuronal excitability. *J. Neurosci.* **19**, 1691–1697 (1999).
15. Kobayashi, K. *et al.* Targeted disruption of the tyrosine hydroxylase locus results in severe catecholamine depletion and perinatal lethality in mice. *J. Biol. Chem.* **270**, 27235–27243 (1995).
16. Lechner, H. A., Lein, E. S. & Callaway, E. M. A genetic method for selective and quickly reversible silencing of mammalian neurons. *J. Neurosci.* **22**, 5287–5290 (2002).
17. Nitabach, M. N., Blau, J. & Holmes, T. C. Electrical silencing of *Drosophila* pacemaker neurons stops the free-running circadian clock. *Cell* **109**, 485–495 (2002).
18. Tan, E. M. *et al.* Selective and quickly reversible inactivation of mammalian neurons *in vivo* using the *Drosophila* allatostatin receptor. *Neuron* **51**, 157–170 (2006).
19. Karpova, A. Y., Tervo, D. G., Gray, N. W. & Svoboda, K. Rapid and reversible chemical inactivation of synaptic transmission in genetically targeted neurons. *Neuron* **48**, 727–735 (2005).
20. Nagel, G. *et al.* Channelrhodopsin-1: a light-gated proton channel in green algae. *Science* **296**, 2395–2398 (2002).
21. Kolbe, M., Besir, H., Essen, L. O. & Oesterhelt, D. Structure of the light-driven chloride pump halorhodopsin at 1.8 Å resolution. *Science* **288**, 1390–1396 (2000).
22. Bamberg, E., Tittor, J. & Oesterhelt, D. Light-driven proton or chloride pumping by halorhodopsin. *Proc. Natl Acad. Sci. USA* **90**, 639–643 (1993).
23. Duschl, A., McCloskey, M. A. & Lanyi, J. K. Functional reconstitution of halorhodopsin. Properties of halorhodopsin-containing proteoliposomes. *J. Biol. Chem.* **263**, 17016–17022 (1988).
24. Hegemann, P., Oesterhelt, D. & Bamberg, E. The transport activity of the light-driven chloride pump halorhoposin is regulated by green and blue light. *Biochim. Biophys. Acta* **819**, 195–205 (1985).
25. Alfonso, A., Grundahl, K., Duerr, J. S., Han, H. P. & Rand, J. B. The *Caenorhabditis elegans unc-17* gene: a putative vesicular acetylcholine transporter. *Science* **261**, 617–619 (1993).
26. Faumont, S. & Lockery, S. R. The awake behaving worm: simultaneous imaging of neuronal activity and behavior in intact animals at millimeter scale. *J. Neurophysiol.* **95**, 1976–1981 (2006).
27. Rosenmund, C. *et al.* Differential control of vesicle priming and short-term plasticity by Munc13 isoforms. *Neuron* **33**, 411–424 (2002).
28. Yoon, H., Enquist, L. W. & Dulac, C. Olfactory inputs to hypothalamic neurons controlling reproduction and fertility. *Cell* **123**, 669–682 (2005).
29. Lu, J., Sherman, D., Devor, M. & Saper, C. B. A putative flip-flop switch for control of REM sleep. *Nature* **441**, 589–594 (2006).
30. Schoppa, N. E. & Westbrook, G. L. Glomerulus-specific synchronization of mitral cells in the olfactory bulb. *Neuron* **31**, 639–651 (2001).
31. Hanks, T. D., Ditterich, J. & Shadlen, M. N. Microstimulation of macaque area LIP affects decision-making in a motion discrimination task. *Nature Neurosci.* **9**, 682–689 (2006).
32. Kanold, P. O., Kara, P., Reid, R. C. & Shatz, C. J. Role of subplate neurons in functional maturation of visual cortical columns. *Science* **301**, 521–525 (2003).
33. Jensen, O. & Lisman, J. E. Novel lists of 7 ± 2 known items can be reliably stored in an oscillatory short-term memory network: interaction with long-term memory. *Learn. Mem.* **3**, 257–263 (1996).
34. Lima, S. Q. & Miesenbock, G. Remote control of behavior through genetically targeted photostimulation of neurons. *Cell* **121**, 141–152 (2005).
35. Shoham, S., O'Connor, D. H., Sarkisov, D. V. & Wang, S. S. Rapid neurotransmitter uncaging in spatially defined patterns. *Nature Methods* **2**, 837–843 (2005).
36. Helmchen, F., Fee, M. S., Tank, D. W. & Denk, W. A miniature head-mounted two-photon microscope. high-resolution brain imaging in freely moving animals. *Neuron* **31**, 903–912 (2001).
37. Pettit, D. L. & Augustine, G. J. Distribution of functional glutamate and GABA receptors on hippocampal pyramidal cells and interneurons. *J. Neurophysiol.* **84**, 28–38 (2000).
38. Volgraf, M. *et al.* Allosteric control of an ionotropic glutamate receptor with an optical switch. *Nature Chem. Biol.* **2**, 47–52 (2006).
39. Chambers, J. J., Banghart, M. R., Trauner, D. & Kramer, R. H. Light-induced depolarization of neurons using a modified Shaker K+ channel and a molecular photoswitch. J. Neurophysiol. **96**, 2792–2796 (2006).
40. Zemelman, B. V., Lee, G. A., Ng, M. & Miesenbock, G. Selective photostimulation of genetically chARGed neurons. *Neuron* **33**, 15–22 (2002).
41. Wang, S. S., Khiroug, L. & Augustine, G. J. Quantification of spread of cerebellar long-term depression with chemical two-photon uncaging of glutamate. *Proc. Natl Acad. Sci. USA* **97**, 8635–8640 (2000).
42. Adams, S. R. & Tsien, R. Y. Controlling cell chemistry with caged compounds. *Annu. Rev. Physiol.* **55**, 755–784 (1993).
43. Dittgen, T. *et al.* Lentivirus-based genetic manipulations of cortical neurons and their optical and electrophysiological monitoring *in vivo*. *Proc. Natl Acad. Sci. USA* **101**, 18206–18211 (2004).
44. MacLean, J. N., Fenstermaker, V., Watson, B. O. & Yuste, R. A visual thalamocortical slice. *Nature Methods* **3**, 129–134 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. Huguenard for discussions; L. Meltzer and H.-C. Tsai for assistance with confocal imaging; R. Airan for statistical assistance; V. Gradinaru for help with calcium imaging; M. Engelhard, J. Rand, D. Oesterhelt, R. Abele, and B. Bauer for plasmids; and K. Zehl and E. Grabski for expert technical assistance. F.Z. is supported by a fellowship from the NIH. L.-P.W. is supported by a fellowship from the California Institute of Regenerative Medicine. F.Z. and L.-P.W. are co-first authors. E.B. and G.N. are supported by grants from the Max-Planck-Society and by the Deutsche Forschungsgemeinschaft. A.G. is supported by grants from the Hessisches Ministerium für Wissenschaft und Kunst, and by the Deutsche Forschungsgemeinschaft. K.D. is supported by NARSAD, APIRE and the Snyder, Culpeper, Coulter, Klingenstein, Whitehall, McKnight, and Albert Yu and Mary Bechmann Foundations, as well as by NIMH, NIDA, and the NIH Director's Pioneer Award Program.

Author Information The GenBank accession number is EF474018 for the 'mammalianized' NpHR sequence and EF474017 for the 'mammalianized' ChR2(1-315) sequence. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests.
Correspondence and requests for materials should be addressed to K.D. (deissero@stanford.edu).

METHODS

Oocyte microinjection and electrophysiology. NpHR complementary RNA was generated using the T7-cap scribe kit from Ambion. Stage V/VI oocytes were prepared as described elsewhere[45]. Each oocyte was injected with 30 to 50 ng cRNA, and incubated for 4 to 7 days at 16 to 18 °C with 1 μM ATR in the medium (90 mM NaCl, 2 mM KCl, 1 mM $MgCl_2$, 1.8 mM $CaCl_2$, 5 mM HEPES, pH 7.4/NaOH) to reconstitute functional halorhodopsins. As a control uninjected oocytes were incubated in the same medium. Oocytes were recorded using two-electrode voltage-clamp (Turbo Tec-05) and illuminated with a continuous He-Ne laser (594 nm, LYHR-0600M, Laser 2000). The maximum light intensity was 3 mW mm^{-2} and was focused to a diameter close to the dimensions of the oocyte. In giant patch experiments from halorhodopsin-expressing oocytes we used a continuous He-Ne laser of 633 nm with light intensities up to 400 mW mm^{-2}.

Lentiviral vector construction. Lentiviral vectors containing *Synapsin I-ChR2–mCherry*, *CaMKIIα-ChR2–mCherry*, and *CaMKIIα-NpHR–EYFP* were based on the FCK(1.3)GW plasmid (gift of P. Osten[43]). For the construction of these lentiviral vectors, the promoter was PCR amplified and cloned into the *Pac*I and *Age*I restriction sites. The transgene *ChR2–mCherry* or *NpHR–EYFP* were PCR amplified and cloned into the *Age*I and *Eco*RI restriction sites. The pLEHYT vector is constructed in the same way as pLECYT[2] by inserting the *NpHR–EYFP* gene into the *Afe*I and *Spe*I restriction sites of pLEGT (gift of E. Wexler and T. Palmer). For both *NpHR–EYFP* and *ChR2–mCherry*, the protein fusion was made using a *Not*I restriction site. The linker between the two proteins is 5′-GCGGCCGCC-3′. The start codon on the fluorescent protein was removed deliberately to avoid translation of the fluorescent protein alone. In addition to the promoter, each lentiviral vector contains the HIV-1 central polypurine tract (cPPT)[46] and the Woodchuck Hepatitis Virus Post-transcriptional Regulatory Element (WPRE)[47] to improve transduction efficiency.

Lentiviral production and transduction. High-titre lentiviruses were produced using a second generation lentiviral system, by co-transfection of 293FT cells (Invitrogen) with pCMVΔR8.74 and pMD2.G (both gifts of M. Sena-Esteves) in addition to the viral vector. A modified version of the protocol presented in Sena-Esteves *et al.*[48] was used. The protocol is reproduced below.

Day 0: Split 4 T-225 flasks (Nunc) of 95% confluent 293FT cells into 1 4-layer CellFactory (Nunc). Culture using 500 ml of DMEM with 10% FBS. Incubate the plates at 37 °C overnight. The cells should reach 90% confluence in 24 h.

Day 1: Perform calcium phosphate transfection. Make DNA mixture containing 690 μg of the viral vector, 690 μg of pCMVΔR8.74 and 460 μg of pMD2.G. Add 5.7 ml of 2 M $CaCl_2$ to the DNA mixture and bring the total volume to 23.75 ml with distilled H_2O. Then, quickly combine the DNA/$CaCl_2$ solution with 23.75 ml of 2× HBS (50 mM HEPES, 1.5 mM Na_2HPO_4, 180 mM NaCl, pH 7.05; note that the pH is important). After quickly mixing by inverting 5 times, add the DNA/$CaCl_2$/HBS solution to 500 ml of room-temperature DMEM with 10% FBS to make the transfection media. Then, exchange the media in the CellFactory with the transfection media.

Day 2: 15 h from the time of transfection, remove the transfection media from the CellFactory and wash the cells 3 times with fresh room-temperature DMEM. Incubation longer than 15 h may lead to cell death and reduced viral titre. Finally, replace the media with 500 ml of fresh DMEM containing 10% FBS and incubate in a 37 °C incubator for 9 h.

Day 2.5: 24 h from the time of transfection, remove the media from the CellFactory and replace with 200 ml of serum-free media (UltraCULTURE, Cambrex) containing 5 mM sodium butyrate. Return the CellFactory to the incubator.

Day 3: 40 h from the time of transfection, collect the 200 ml of media from the CellFactory. This is the viral-containing supernatant. Centrifuge at 1,000 r.p.m. for 5 min to precipitate large cell debris and then filter the viral supernatant using a 0.45 μm low-protein-binding filter flask. Then, centrifuge the supernatant using a SW-28 rotor (Beckman Coulter) for 2 h at 55,000*g* to precipitate the virus. Usually 6 centrifuge tubes are required to concentrate all of the viral supernatant. Before spinning, add 2 ml of PBS containing 20% sucrose to the bottom of the centrifuge tube to remove any remaining cell debris during centrifugation. After centrifugation, gently decant the liquid from the centrifuge tubes and resuspend all 6 viral pellets with 100 μl of 4 °C PBS. Then, aliquot the viral solution and store at −80 °C for future use. If desired, 10 ml of unconcentrated viral supernatant can be stored before centrifugation for *in vitro* use in cultured neurons. For culture applications, neurons can be transduced simply by adding 50 μl of unconcentrated viral supernatant per 24-well plate well. Protein expression can be observed 4 to 5 days later. For *in vivo* applications, concentrated virus can be directly injected into the mammalian brain as detailed below.

Stereotactic injection in the mouse brain. Female C57BL/6 mice, 5–7 weeks old, were housed according to the Laboratory Vertebrate Animals protocols at Stanford. The virus solution was delivered into the CA3 region of the hippocampus through stereotactic injection. Surgeries were performed under aseptic conditions. For anaesthesia, ketamine (16 mg kg^{-1} of body weight) and xylazine (5 mg kg^{-1} of body weight) cocktail were injected intraperitoneally. Fur was sheared from the top of the animal's head and the head was placed in a stereotactic apparatus (David Kopf Instruments). A midline scalp incision was made and a 1-mm-diameter craniotomy was drilled (anteroposterior, −2.0 mm from bregma; lateral, 2 mm; ventral, 2.2 mm). A glass micropipette was refilled with 3.0 μl of concentrated lentivirus solution using a programmable pump (PHD 2000, Harvard Apparatus) and 1 μl of lentivirus solution was injected at each site at a concentration of 0.1 μl min^{-1}.

Electrophysiology. Patch-clamp recordings in oocytes[4,20] and neurons[2] were carried out as previously described. For whole-cell and cell-attached recording in cultured hippocampal neurons or acute brain slices, three intracellular solutions containing chloride were prepared: 4 mM chloride (135 mM K-gluconate, 10 mM HEPES, 4 mM KCl, 4 mM MgATP, 0.3 mM Na_3GTP, titrated to pH 7.2); 10 mM chloride (129 mM K-gluconate, 10 mM HEPES, 10 mM KCl, 4 mM MgATP, 0.3 mM Na_3GTP, titrated to pH 7.2); or 25 mM chloride (114 mM K-gluconate, 10 mM HEPES, 25 mM KCl, 4 mM MgATP, 0.3 mM Na_3GTP, titrated to pH 7.2). For cultured hippocampal neurons, Tyrode's solution was employed as the extracellular solution (125 mM NaCl, 2 mM KCl, 3 mM $CaCl_2$, 1 mM $MgCl_2$, 30 mM glucose, and 25 mM HEPES, titrated to pH 7.3). For preparation of acute brain slices, mice were killed 2 weeks after viral injection. Acute brain slices (250 μm) were prepared in ice-cold cutting solution (64 mM NaCl, 25 mM NaHCO3, 10 mM glucose, 120 mM sucrose, 2.5 mM KCl, 1.25 mM NaH_2PO_4, 0.5 mM $CaCl_2$, 7 mM $MgCl_2$, and equilibrated with 95% O_2/5% CO_2) using a vibratome (VT1000S, Leica). Slices were incubated in oxygenated ACSF (124 mM NaCl, 3 mM KCl, 26 mM $NaHCO_3$, 1.25 mM NaH_2PO_4, 2.4 mM $CaCl_2$, 1.3 mM $MgCl_2$, 10 mM glucose, and equilibrated with 95% +O_2/5% CO_2) at 32 °C for 30 min to recover.

Fura-2-AM loading and Ca^{2+} imaging. For calcium imaging, lentiviruses were injected into the cortex of C57BL/6 mice at postnatal day 4 or 5 and acute brain slices were prepared 7 to 8 days later as described for adult mice. Fura-2-AM was loaded into the regions of interest as previously described[44]. Calcium activity was recorded on an upright fluorescence microscope (DM LFSA, Leica) with a ×20, 0.5 numerical aperture water immersion objective. The GFP/mRFP1 dual-bandpass dichroic (51022bs, Chroma) and emitter (51022m, Chroma) was used for both photostimulation and Ca^{2+} imaging. NpHR was excited using the 593/40 bandpass filter (FF01-593/40-25, Semrock) and ChR2 was excited using the 472/30 bandpass filter (FF01-472/30-25, Semrock). Simultaneous activation of NpHR and ChR2 was achieved using either the GFP/mRFP1 dual bandpass filter (51022x, Chroma) or the combination of a 473 nm blue diode laser (Crystal Lasers) and a 300 W Lambda DG-4 with a 593 nm excitation filter. Fura-2 was excited using the 340/26 bandpass filter (FF01-340/26-25, Semrock). Frames were collected using a cooled CCD camera (Retiga EXi, QImaging) at 0.5 Hz using MetaFluor (Molecular Devices). $[Ca^{2+}]_i$ changes were detected in single cells and expressed as the fluorescence intensity ratios $\Delta F/F_0$ with the resting fluorescence value F_0 determined at the beginning of each experiment.

Transgenic *C. elegans* lines and transgenes. The *NpHR* gene was placed under the muscle-specific *myo-3* promoter (untagged NpHR in transgene *zxEx29[pmyo-3-NpHR; lin-15$^+$]* and NpHR–ECFP in transgene *zxEx30[pmyo-3-NpHR–ECFP; rol-6d]*) or under the cholinergic motor neuron specific *unc-17* promoter (NpHR–ECFP in transgene *zxEx33[punc-17-NpHR –ECFP; lin-15$^+$]*). The NpHR–ECFP fusion (*zxEx30*, and *zxEx34*, see below) was employed to assess expression pattern. NpHR–ECFP (*zxEx30*) animals showed light-induced effects that were comparable to the untagged version (*zxEx29*).

For co-activation of ChR2/NpHR in muscles or cholinergic motor neurons, transgenes *zxEx32[pmyo-3-NpHR; pmyo-3-ChR2(H134R) –EYFP; lin-15$^+$]* and *zxEx34[punc-17-NpHR –ECFP; punc-17-ChR2(H134R)–YFP; rol-6d]* were used. A list of the transgenes and worm lines can be found in Supplementary Table 1.

***C. elegans* stimulation and electrophysiology.** *C. elegans* were grown in the presence of ATR, as described previously[6] except that the ATR concentration was reduced to 5 μM. For NpHR activation in behavioural assays, we used a Zeiss Axiovert 40 with a 50 W HBO mercury lamp, HQ-F41-007 excitation/emission filter set (excitation 530–560 nm, AHF Analysentechnik), or Zeiss Filterset 00 (excitation 530–585 nm), and either a 10× or 20× objective (10.2 and 4.4 mW mm^{-2}, respectively). For co-activation of ChR2, a blue DPSS laser (473 nm at 7 mW mm^{-2}, Pusch OptoTech) was used. Patch clamping of *C. elegans* muscle was performed as previously described[6], at a holding potential of −40 mV, using a Zeiss Axioskop 2 FS plus, equipped with a 100 W HBO mercury lamp, HQ-F41-007 filter set, and a ×40 water immersion objective (6.7 mW mm^{-2}).

45. Grygorczyk, R., Hanke-Baier, P., Schwarz, W. & Passow, H. Measurement of erythroid band 3 protein-mediated anion transport in mRNA-injected oocytes of *Xenopus laevis*. *Methods Enzymol.* **173**, 453–466 (1989).

46. Zennou, V. *et al.* The HIV-1 DNA flap stimulates HIV vector-mediated cell transduction in the brain. *Nature Biotechnol.* **19,** 446–450 (2001).
47. Brun, S., Faucon-Biguet, N. & Mallet, J. Optimization of transgene expression at the posttranscriptional level in neural cells: implications for gene therapy. *Mol. Ther.* **7,** 782–789 (2003).
48. Sena-Esteves, M., Tebbets, J. C., Steffens, S., Crombleholme, T. & Flake, A. W. Optimized large-scale production of high titre lentivirus vector pseudotypes. *J. Virol. Methods* **122,** 131–139 (2004).

Vol 446|5 April 2007|doi:10.1038/nature05731

ARTICLES

Mechanism of auxin perception by the TIR1 ubiquitin ligase

Xu Tan[1], Luz Irina A. Calderon-Villalobos[2], Michal Sharon[3], Changxue Zheng[1], Carol V. Robinson[3], Mark Estelle[2] & Ning Zheng[1]

Auxin is a pivotal plant hormone that controls many aspects of plant growth and development. Perceived by a small family of F-box proteins including transport inhibitor response 1 (TIR1), auxin regulates gene expression by promoting SCF ubiquitin-ligase-catalysed degradation of the Aux/IAA transcription repressors, but how the TIR1 F-box protein senses and becomes activated by auxin remains unclear. Here we present the crystal structures of the *Arabidopsis* TIR1–ASK1 complex, free and in complexes with three different auxin compounds and an Aux/IAA substrate peptide. These structures show that the leucine-rich repeat domain of TIR1 contains an unexpected inositol hexakisphosphate co-factor and recognizes auxin and the Aux/IAA polypeptide substrate through a single surface pocket. Anchored to the base of the TIR1 pocket, auxin binds to a partially promiscuous site, which can also accommodate various auxin analogues. Docked on top of auxin, the Aux/IAA substrate peptide occupies the rest of the TIR1 pocket and completely encloses the hormone-binding site. By filling in a hydrophobic cavity at the protein interface, auxin enhances the TIR1–substrate interactions by acting as a 'molecular glue'. Our results establish the first structural model of a plant hormone receptor.

Auxin has long been recognized as an important phytohormone that regulates plant growth in response to diverse developmental and environmental cues[1,2]. The major naturally occurring auxin, indole-3-acetic acid (IAA), coordinates many plant growth processes by modulating gene expression, which leads to changes in cell division, elongation and differentiation. How the auxin signal is perceived and interpreted by plant cells has been a central question in plant biology.

Recent genetic and molecular studies in *Arabidopsis* have revealed a crucial intracellular auxin signalling pathway in which a ubiquitin-dependent proteolytic system has a key role in sensing and transducing the hormone signal to transcriptional programs[3]. At the centre of the signalling cascade is the ubiquitin-ligase complex, SCF^{TIR1}, which promotes the ubiquitin-dependent proteolysis of a family of transcriptional regulators known as Aux/IAAs in an auxin-dependent manner[4]. Degradation of the Aux/IAAs activates the auxin response factor (ARF) family of transcription factors, whose activities in regulating auxin responsive genes are otherwise inhibited by the Aux/IAA proteins[5–9]. The *Arabidopsis* genome encodes 29 Aux/IAA proteins[10–12], most of which share a highly conserved degron sequence (called domain II) that can be recognized by the SCF^{TIR1} complex[13–15]. Importantly, two recent studies have revealed that TIR1, the F-box protein subunit of SCF^{TIR1}, functions as a true auxin receptor[16,17]. It has been shown that auxin binds directly to SCF^{TIR1} and promotes the interaction between TIR1 and Aux/IAAs. In *Arabidopsis*, TIR1 and its closest paralogues AFB1 to AFB5 belong to the C3 subfamily of leucine-rich-repeat-containing F-box proteins[18]. Together with TIR1, AFB1 to AFB3 have been found to function as redundant auxin receptors, collectively mediating auxin-regulated responses throughout plant growth and development[19].

Besides IAA, several naturally occurring and many synthetic auxins have been identified. These compounds are chemically diverse, sharing only a planar unsaturated ring and a side chain with a carboxyl group[20]. Despite several hypotheses on the determinants of auxin activity[21,22], how these variable chemical substances elicit a common phytohormone function remains unclear. In addition to IAA, TIR1 has been reported to recognize at least two other synthetic auxin analogues, 1-naphthalene acetic acid (1-NAA) and 2,4-dichlorophenoxyacetic acid (2,4-D)[16,17]. Similarly to IAA, both compounds are able to promote the binding of Aux/IAA proteins to the TIR1 F-box protein. A structural explanation of how TIR1 perceives and is in turn activated by these structurally distinct auxins will help to elucidate the structure–activity relationships of auxins.

Featuring interchangeable substrate receptor subunits, the SCF and SCF-like cullin-RING complexes control a broad spectrum of cellular functions in eukaryotes by ubiquitinating diverse protein substrates[23]. Substrate recognition by known cullin-RING ubiquitin ligases often involves post-translational modification of the substrate, for instance, serine phosphorylation and proline hydroxylation. Auxin-mediated SCF^{TIR1}–substrate interaction represents a new mechanism in which a naturally occurring small molecule directly stimulates the substrate-binding activity of a cullin-RING ubiquitin ligase. The structural basis of this ubiquitin-ligase regulatory mechanism and its implications beyond phytohormone signalling await clarification.

Here we report a series of crystallographic studies of auxin perception and auxin-promoted substrate-binding of the auxin receptor TIR1. Our results reveal the structural mechanism underlying auxin perception by TIR1, suggest a potential role of inositol hexakisphosphate ($InsP_6$) in plant hormone signalling, and indicate a new direction for developing therapeutic drugs that target human ubiquitin ligases.

Auxin enhances TIR1–Aux/IAA interaction

We overexpressed and purified full-length *Arabidopsis* TIR1 in complex with the full-length SCF^{TIR1} adaptor, ASK1, using the baculovirus–insect cell system. Auxin-enhanced TIR1–Aux/IAA interaction was readily detected in a native gel shift assay. Consistent with early studies, we also observed an intrinsic low affinity between TIR1 and the

[1]Department of Pharmacology, University of Washington, School of Medicine, Box 357280, Seattle, Washington 98195, USA. [2]Department of Biology, Indiana University, Bloomington, Indiana 47405, USA. [3]Department of Chemistry, University of Cambridge, Cambridge, CB2 1EW, UK.

bacteria-produced IAA7 protein in the absence of auxin (Supplementary Fig. 1)[16,17]. These results confirm that auxin regulates SCF^{TIR1} by directly promoting the interactions between TIR1 and its polypeptide substrates[16,17].

To increase the substrate affinity of TIR1, auxin might either bind to an allosteric site and induce conformational changes of the F-box protein, or interact with both TIR1 and the Aux/IAA substrate and extend the binding interface between the two polypeptides. To differentiate between these two distinct mechanisms and to reveal how different auxins can bind to and regulate TIR1, we have determined the crystal structures of the *Arabidopsis* TIR1–ASK1 complex alone and in complexes with three auxin compounds as well as an Aux/IAA degron peptide derived from the IAA7 protein, together and separately at resolutions between 1.8 Å and 2.5 Å (Supplementary Table 1).

Overview of the structure

The TIR1–ASK1 complex has a mushroom-shaped structure with the leucine-rich-repeat (LRR) domain of TIR1 being the 'cap' and the F-box motif of TIR1 together with ASK1 forming the 'stem' (Fig. 1a). The spatial arrangement of the domains in the complex is similar to that seen in the human Skp2–Skp1 complex, in which the F-box protein Skp2 also features an LRR domain[24]. When modelled onto the human SCF complex[25], the TIR1–ASK1 complex orients the top surface of the mushroom cap towards the RING subunit of the E3 machinery (Supplementary Fig. 2). With twice as many LRR motifs as Skp2, the LRR domain of TIR1 (TIR1-LRR) folds into a twisted horseshoe-shaped solenoid (Fig. 1). Strikingly, an unexpected $InsP_6$ molecule is found near the centre of the TIR1-LRR fold in close vicinity to the auxin-binding site (Fig. 1).

The TIR1-LRR domain is solely responsible for auxin perception and IAA7 substrate recruitment. A single top surface pocket in TIR1-LRR performs both functions (Fig. 1). Auxin is recognized at the bottom of the pocket through an overall hydrophobic and partially promiscuous binding site that can accommodate variable planar ring structures. A key TIR1 arginine residue (Arg 403) is supported by the nearby $InsP_6$ cofactor and selects the carboxyl group shared by most auxins. Immediately above the auxin-binding site, the highly coiled IAA7 peptide is docked to the upper part of the surface pocket, packing its central hydrophobic consensus motif directly against auxin and completely covering the plant hormone (Fig. 1a). A superposition analysis shows that the structure of the TIR1 LRR domain remains largely unchanged with or without auxins and/or the IAA7 peptide bound.

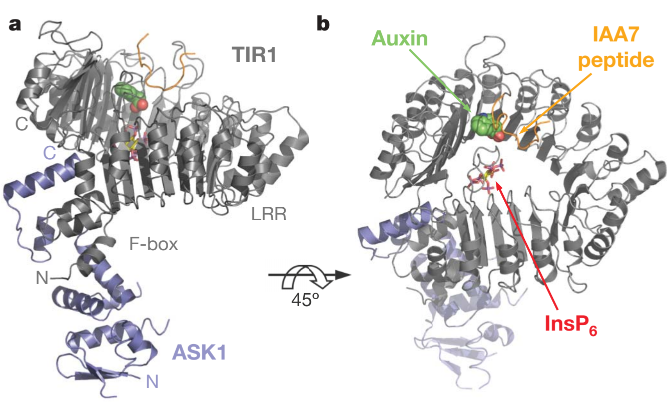


Figure 1 | Crystal structure of the TIR1–ASK1 complex with auxin and the IAA7 degron peptide. **a**, **b**, Two views of the complex structure are shown as a ribbon diagram. TIR1, ASK1 and the IAA7 substrate peptide are coloured grey, blue and orange, respectively. The F-box and LRR domains of TIR1 are labelled. Auxin is represented by a space-filling model (CPK). The $InsP_6$ molecule is shown as a stick model.

Architecture of the TIR1-LRR domain

The TIR1-LRR domain consists of 18 LRRs and a carboxy-terminal cap sequence and adopts an overall architecture that is distinct from other known LRR structures[26]. Each TIR1-LRR contains a β-strand followed by an α-helix. Packing in tandem, the TIR1-LRRs assemble into the expected solenoid structure with an overall horseshoe-like shape. The highly curved solenoid fold is characterized by a concave surface formed mostly by a parallel β-sheet and a convex surface lined with α-helices (Fig. 2a, b). Unlike the regular horseshoe-shaped structures reported for other LRR proteins[27,28], however, the TIR1-LRR solenoid has an abrupt kink in the middle, highlighted by the α-helix of the eighth LRR protruding from the convex surface (Fig. 2a, b). The two halves of the TIR1-LRR horseshoe are packed at an angle and the ends of the horseshoe are close to each other (Fig. 2b). The C-terminal cap sequence, which extends the LRRs by three antiparallel β-strands, is positioned right above the first LRR and essentially closes the curved solenoid as a complete circle. The resulting twisted structure can be likened to a thick 'spring washer', which is overall a one-coil helix (Fig. 2b).

The unique architecture of the TIR1-LRR domain is critical for the function of the auxin receptor. The intra-repeat loops associated with LRRs 2, 12 and 14 are unusually long and located at the top surface of the TIR1-LRR solenoid (Fig. 2). Of the three long loops, the LRR-2 loop (loop-2) plays a pivotal role in constructing the auxin- and substrate-binding surface pocket by interacting with the nearby concave surface of the TIR1-LRR solenoid. This part of the solenoid inner wall is mainly formed by the C-terminal cap sequence and is brought to the close vicinity of the second repeat by the twist in the middle of TIR1-LRR (Fig. 2). With its flanking regions held by the C-terminal cap β-sheet, the central part of loop-2 forms a 3_{10}-helix and approaches the concave surface and the other two long loops at the opposite side of the solenoid. Together, the three long loops and the nearby concave surface of the twisted LRR fold form the auxin- and substrate-binding pocket (Fig. 2c).

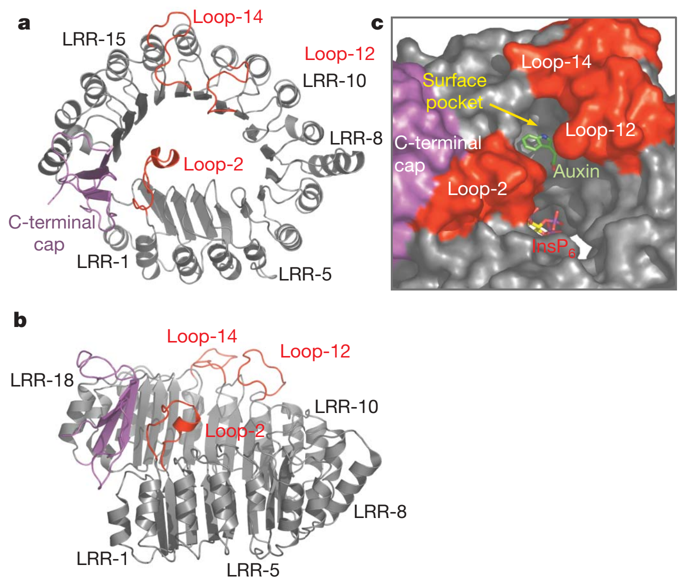


Figure 2 | Architecture of the TIR1-LRR domain. **a**, **b**, Two orthogonal views of the LRR domain of TIR1. The eighteen LRR motifs are numbered and shown in grey. The C-terminal cap sequence is shown in magenta. The three substrate-contacting top surface loops in LRR-2, LRR-12 and LRR-14 are shown in red and labelled as loop-2, loop-12 and loop-14. **c**, Surface representation of the TIR1-LRR domain, focusing on the auxin- and substrate-binding pocket created by the intra-repeat loop of LRR-2 (loop-2), which packs against the concave surface of the LRR solenoid. The TIR1-bound auxin and the nearby $InsP_6$ are represented as stick models.

Identification of $InsP_6$ in TIR1

In addition to an unusual architecture, the TIR1-LRR domain is distinct from other LRR proteins in having a previously unknown co-factor tightly bound near the centre of the solenoid fold. Crystallographic analysis of the TIR1–ASK1 complex at 1.8 Å resolution reveals an island of unexplained electron density under the 3_{10}-helix of loop-2. Surrounded by numerous residues on the concave surface of the TIR1-LRR domain, the extra density is characterized by six similarly sized branches, all connected to a central body (Fig. 3a).

The distinctive features of the unexpected electron density strongly suggest that it belongs to an $InsP_6$ molecule co-purified with the TIR1–ASK1 protein complex from insect cells. A myo-inositol-1,2,3,4,5,6-hexakisphosphate molecule would have a well matched electron density pattern and this would also explain its close interactions with a highly basic surface area of TIR1 (Fig. 3b), which consists of ten positively charged TIR1 residues (Fig. 3a). In addition, mass spectrometry analysis shows that the observed molecular weight of the unknown molecule is essentially identical to the calculated molecular weight of $InsP_6$, unequivocally confirming its chemical identity (Supplementary Fig. 3). The properties of the inositol-hexakisphosphate-binding site and its spatial relationship to the nearby auxin-binding site in TIR1 suggest that $InsP_6$ is probably a structural co-factor of the auxin receptor (see below).

The TIR1 auxin-binding pocket

As revealed in both structures of TIR1–IAA and TIR1–IAA–Aux/IAA degron peptide, TIR1 recognizes auxin through the bottom portion of the surface pocket formed between the LRR-2 loop and the solenoid inner surface. Facing towards the central channel of the LRR solenoid, the pocket is analogous to a three-walled room with an open ceiling (Fig. 2c, 4a). The 3_{10}-helix of loop-2 forms one 'wall', while the highly curved concave surface presented by LRR-12 to LRR-16 defines the other two. The floor of the pocket is made of a layer of residues from both loop-2 and the concave surface (His 78, Arg 403, Ser 438, Ser 462 and Glu 487), together with a water molecule sequestered in the middle (Fig. 4a). Below the floor, another layer of TIR1 residues (Arg 436, Met 460 and Lys 485), all interacting with $InsP_6$, supports the pocket from underneath and leaves it accessible only from the top. The integrity of the pocket floor is crucial for auxin binding: one amino-acid mutation (S462E) is sufficient to abolish auxin-enhanced binding but does not inhibit basal substrate binding by TIR1 (Fig. 4c).

IAA, the natural auxin, binds to the base of the TIR1 pocket via two important functional moieties, the side-chain carboxyl group and the indole ring (Fig. 4a and Supplementary Fig. 4). The carboxyl group of IAA anchors the plant hormone to the bottom of the TIR1 pocket by forming a salt bridge and two hydrogen bonds with two residues from the pocket floor (Arg 403 and Ser 438). Meanwhile, the indole ring of IAA stacks on top of the pocket floor with its edge packing against the surrounding walls through hydrophobic interactions and van der Waals contacts. On the loop-2 side of the pocket, the benzene region of the auxin indole ring interacts with two TIR1 phenylalanine residues (Phe 79 and Phe 82). On the concave surface side, the rest of the auxin indole ring is partially sandwiched between two parallel layers of TIR1 residues and is therefore mainly in contact with the TIR1 polypeptide backbone (Fig. 4b). Unique to IAA, a hydrogen bond is donated by the NH group of the auxin indole ring to a nearby carbonyl group, which belongs to the backbone of a slightly twisted TIR1 β-strand (Fig. 4a, d).

Despite having different ring structures, the two synthetic auxin analogues 1-NAA and 2,4-D bind to TIR1 in a manner similar to that of IAA (Fig. 4d). Through a common carboxyl group, both auxin analogues are tethered to the floor of the TIR1 pocket. The naphthalene ring of 1-NAA, which is slightly larger than an indole ring, is placed in the cavity with an orientation almost identical to the auxin indole ring. The dichlorophenyl ring of 2,4-D fits into the pocket base with its widest dimension (defined by the two large chlorines) aligned with that of the pocket, thus mimicking the double-rings of the other two auxin compounds (Fig. 4d). In both cases, the ring structure of the compound is accommodated by the overall shape and generally hydrophobic properties of the TIR1 cavity. The auxin-binding site of TIR1 is therefore defined by two highly selective polar residues

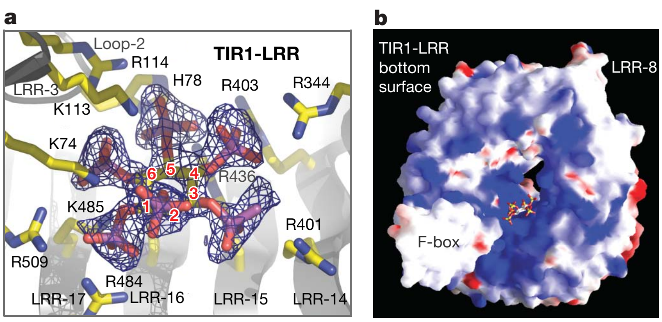


Figure 3 | $InsP_6$ as a TIR1 co-factor. **a**, A close-up view of the TIR1-bound $InsP_6$. The $InsP_6$ molecule is shown as a stick model, together with its positive $F_o - F_c$ electron density calculated and contoured at 4σ before it was built into the TIR1 model. TIR1 residues that are in direct contact with $InsP_6$ are shown as a stick model. **b**, Electrostatic surface potential of the TIR1-LRR bottom surface: positive (blue) to negative (red). The F-box domain and LRR-8 of TIR1 are labelled to indicate the orientation of the domain.

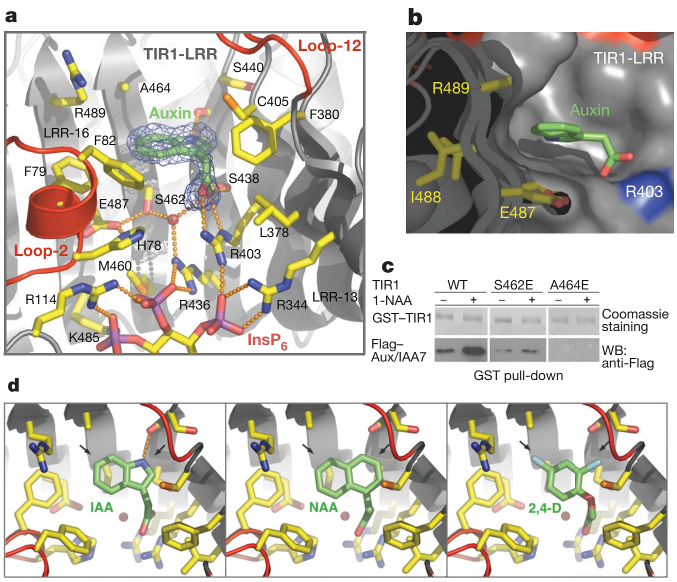


Figure 4 | Recognition of auxin by TIR1. **a**, A close-up view of the auxin-binding site with its nearby $InsP_6$ molecule. The auxin molecule (IAA) is shown as a green stick model, together with its positive $F_o - F_c$ electron density calculated and contoured at 3σ before it was built into the complex model. The TIR1 residues surrounding auxin and right underneath the auxin-binding site are shown as a yellow stick model. A central water molecule as part of the pocket floor is shown as a red sphere. The hydrogen-bond and salt-bridge network connecting auxin and $InsP_6$ are indicated by orange dashed lines. **b**, A slab view of the TIR1 cavity accommodating the auxin indole ring. The TIR1-LRR domain is in surface representation. Three layers of TIR1 residues defining a wall for the auxin-binding site are modelled as sticks. Arg 403 interacts with the auxin carboxyl group and is coloured blue. **c**, GST pull-down assay with (+) or without (−) 50 μM 1-NAA. In the upper panel, purified recombinant wild type (WT) or mutants GST–TIR1 are visualized by Coomassie staining for loading control. In the lower panel, Flag-tagged IAA7 pulled down by GST–TIR1 was detected by western blot. **d**, A top view of the TIR1 cavity with IAA, 1-NAA and 2,4-D bound. The three auxin analogues are all shown as green stick models. In each panel, two black arrows indicate the widest dimension of each compound.

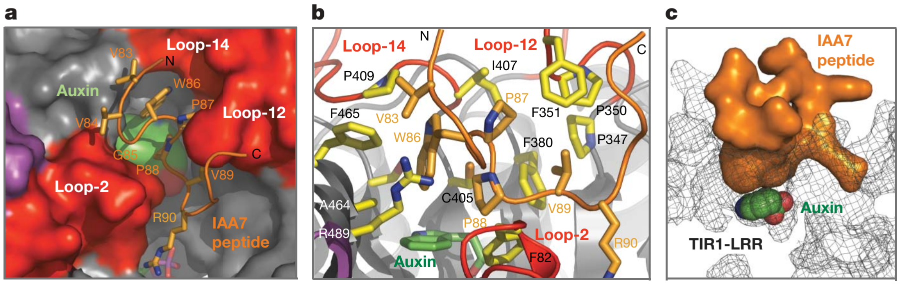


Figure 5 | Binding of an Aux/IAA degron peptide on auxin-bound TIR1. **a**, An overall view of the TIR1 surface pocket occupied by auxin (green) at the bottom and the highly coiled IAA7 peptide (orange) on top. The surface of the three long top surface loops of TIR1 responsible for ligand binding is coloured red. **b**, A close-up side view of the central GWPPV motif in the IAA7 peptide upon binding to TIR1. Interacting residues of the substrate peptide and TIR1 are shown as orange and yellow stick models, respectively. Auxin is shown as a green stick model. To clarify the view, two hydrophobic residues, one from the peptide and the other from loop-2 of TIR1, are not shown. **c**, A slab view of the TIR1–auxin–IAA7 peptide complex, showing that auxin fills a cavity between two proteins. The molecular surface of TIR1 is shown in grey mesh. The IAA7 peptide and auxin are shown in orange surface representation and green CPK, respectively.

(Arg 403 and Ser 438) spatially coupled to a less selective hydrophobic cavity with a fixed shape. This partially promiscuous hormone-binding site in TIR1 explains how the auxin receptor can potentially bind a variety of auxin compounds.

TIR1–substrate peptide interactions

The IAA7 degron peptide has a predominantly hydrophobic sequence and binds to the auxin-bound TIR1 pocket through extensive hydrophobic interactions (Fig. 5a, b). The mouth of the TIR1 pocket is formed mostly by the three extra long LRR loops and is decorated almost exclusively with hydrophobic residues (Fig. 5b). In between the pocket opening and the auxin molecule, another layer of TIR1-LRR residues presents an additional hydrophobic surface. The 13-amino-acid substrate peptide adopts a highly coiled conformation and fills the TIR1 pocket with the central hydrophobic consensus motif GWPPV lying on top of auxin (Fig. 5a, b). The two flanking regions of the motif also contribute to complex formation. The hydrophobic amino-terminal region runs along the pocket rim and the polar C-terminal region curls around the LRR-12 loop of TIR1 (Fig. 5a).

The conserved central GWPPV motif represents the hallmark of the Aux/IAA degron[13]. Genetic screens in *Arabidopsis* have shown that mutations in this motif lead to stabilization of Aux/IAAs and reduced auxin response. Two amino acids in the motif, tryptophan and the second proline, interact with the surrounding hydrophobic wall of the TIR1 pocket and stack against the auxin molecule lying underneath, packing against the auxin indole ring and the auxin side chain, respectively (Fig. 5b and Supplementary Fig. 4). The positions of these two residues are partially maintained by the other proline in the middle, which itself also forms hydrophobic interactions with surrounding TIR1 residues. Similar to the three central residues of the motif, the glycine residue at the first position is invariant among all Aux/IAAs[13]. In the structure, the glycine residue is located at a critical position, where flexibility of the peptide is required for the N-terminal region of the substrate peptide to take a sharp turn and continue interacting with TIR1 (Fig. 5a). Although the valine residue at the end of the motif is not strictly conserved in all Aux/IAA family members, its conserved hydrophobic feature is crucial for its interactions with the nearby hydrophobic residues of TIR1 (Fig. 5b). As a whole, the GWPPV motif of the substrate peptide, together with auxin, nucleates a hydrophobic core networked with the TIR1 hydrophobic surface pocket. When a single non-polar TIR1 residue in the pocket is altered to a charged amino acid (A464E), both the basal and auxin-enhanced binding of the IAA7 substrate to TIR1 are lost (Fig. 4c).

Upon docking to auxin-bound TIR1, the IAA7 peptide almost completely encloses the three-walled TIR1 pocket, covering both its open 'ceiling' and the empty side facing the central channel of the LRR solenoid (Fig. 5a). Thus, the receptor-bound hormone is expected to remain trapped in the pocket until the substrate polypeptide is released. Overall, even though the Aux/IAA substrate itself does not have any detectable affinity to auxin, an optimal binding site for the hormone is cooperatively formed by the substrate polypeptide and TIR1.

Auxin regulation of TIR1–Aux/IAA binding

The crystal structure of TIR1–ASK1 in the ligand-free and different ligand-bound forms reveals that auxin is not an allosteric regulator of SCF^{TIR1}. Superposition analysis of the TIR1–ASK1 structures in the absence and presence of auxin shows that except for the side-chain rearrangement of three local residues (Supplementary Fig. 5), auxin binding does not induce significant conformational changes of the hormone receptor. Furthermore, the auxin-binding site is disqualified as an allosteric site because it is in the same pocket as the substrate-binding site.

Instead, auxin enhances the substrate-binding activity of TIR1 mainly by filling a cavity between the two proteins, thereby extending the protein-interaction interface. Upon interacting with both TIR1 and the substrate polypeptide, auxin mediates the formation of a continuous hydrophobic core among the three. Thus, we conclude that auxin promotes SCF^{TIR1}–substrate binding by acting as a 'molecular glue' rather than an allosteric switch (Fig. 6). Such a regulatory mechanism is in fact consistent with the partially promiscuous nature of auxin binding to the F-box protein. Because the structural role of auxin is to support the GWPPV motif of the Aux/IAA substrate by providing an overall hydrophobic base at the bottom of the TIR1 surface pocket, moderate variations in auxin structure might be tolerated as long as they can be accommodated by the auxin-binding site and do not interfere with Aux/IAA binding.

$InsP_6$ as a possible co-factor of TIR1

$InsP_6$, also known as "phytate", was first identified in plants and later found in other eukaryotes[29]. Many lines of evidence indicate that the $InsP_6$ revealed in the crystals is a specific functional cofactor of TIR1.

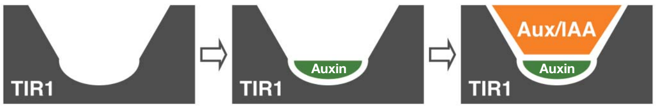


Figure 6 | A model of auxin-regulated TIR1–substrate interactions. A schematic diagram of auxin functioning as a 'molecular glue' to enhance TIR1–substrate interactions. In contrast to an allosteric mechanism, auxin binds to the same TIR1 pocket that docks the Aux/IAA substrate. Without inducing significant conformational changes in its receptor, auxin increases the affinity of two proteins by simultaneously interacting with both in a cavity at the protein interface.

First, the robust co-purification of $InsP_6$ with the TIR1–ASK1 complex through multiple steps of purification shows that the $InsP_6$ is tightly associated with TIR1. Second, eight out of the ten positively charged TIR1 residues that interact with $InsP_6$ are strictly conserved among the *Arabidopsis* AFBs (Supplementary Figs 6 and 7), suggesting that $InsP_6$-binding is essential for this specific subfamily of F-box proteins. Third, the $InsP_6$ molecule interacts with several structural elements of TIR1 that are functionally important. These include both halves of the TIR1-LRR solenoids, loop-2 and Arg403 (Fig. 1b and 3a).

Our discovery of the presence of $InsP_6$ in TIR1 represents the second time $InsP_6$ has been identified in a protein structure via X-ray crystallography[30]. On the basis of a number of other studies, $InsP_6$ has now been shown to regulate a variety of cellular functions such as DNA repair, chromatin remodelling, endocytosis and nuclear messenger RNA export[31–34]. Although the essential role of $InsP_6$ in the *Arabidopsis* auxin receptor TIR1 remains to be confirmed, future studies might elucidate how auxin signalling potentially interfaces with other plant signalling events that govern $InsP_6$ metabolism.

Biological and pharmacological implications

In plants, multiple phytohormone signalling pathways are now known to be regulated by ubiquitin ligases[35]. In particular, jasmonic acid signalling requires COI1, an F-box protein with high sequence similarity to TIR1[36]. Our structural results predict that COI1 adopts a TIR1-like structure and possibly functions as a jasmonic acid receptor. Most of the auxin-contacting residues in TIR1 are indeed not conserved in COI1 (Supplementary Fig. 6). Although TIR1 orthologues so far are only found in plants, a small ligand-sensing site regulating substrate recruitment could conceivably be evolved in a different structural context in other eukaryotic ubiquitin ligases.

An increasing number of human disorders has now been associated with defective ubiquitin-ligase–substrate interactions owing to mutations of the ligases themselves, ubiquitination substrates or upstream signalling proteins responsible for substrate priming[37]. The regulatory mechanism of TIR1 by auxin suggests that it is possible for small molecules to promote protein–protein interactions in ubiquitin ligases. Developing protein interaction agonists could potentially resurrect the disease-causing defective ubiquitination processes. Further studies following this principle may open up the development of pharmaceutical agents targeting ubiquitin ligases and potentially other protein interaction systems that are impaired by genetic alterations.

METHODS

Protein preparation and GST pull-down assay. The full-length *Arabidopsis thaliana* TIR1 and ASK1 were co-expressed as a glutathione S-transferase (GST)-fusion protein and a 6X His-tag fusion protein, respectively, in Hi5 suspension insect cells. The TIR1–ASK1 complex was isolated from the soluble cell lysate by glutathione affinity chromatography. After cleavage by TEV, the complex was further purified by anion exchange and gel filtration chromatography and concentrated by ultrafiltration to 4 mg ml^{-1}. GST pull-down assays were carried out using affinity-purified recombinant GST–TIR1–ASK1 (insect cells) and His-Flag-tagged IAA7 proteins (*Escherichia coli*).

Crystallization and data collection. The TIR1–ASK1 crystals were grown at 25 °C by the hanging-drop vapour-diffusion method with 1.5 µl protein samples mixed with an equal volume of reservoir solution containing 100 mM BTP (pH = 6.0), 10–14% PEG 20,000, 200 mM NaCl and 5 mM DTT. Diffraction-quality crystals were obtained by macro-seeding at 4 °C. The crystals form in space group $C2$ ($a = 102.5$ Å, $b = 80.5$ Å, $c = 125.0$ Å, $\alpha = \gamma = 90°$, $\beta = 105.1°$) and contain one molecule in the asymmetric unit. The TIR1–ASK1 derivative crystals were prepared by soaking the native crystal in the reservoir solution supplemented with 16 compounds containing Hg, Au and Pt for 2 h. Among them, five compounds (thimerasol, $(NH_4)_2PtCl_4$, K_2PtCl_6, K_2PtCl_4 and $KAu(CN)_2$) produced diffracting crystals that are isomorphous to the native crystals and were used for data collection. The TIR1–ASK1 crystals were also soaked with different auxins (IAA, 2,4-D and NAA, from Sigma) and an Aux/IAA peptide (NH2-QVVGWPPVRNYRK-COOH) to obtain complex crystals (see Supplementary Table 1). The data sets were collected at the BL8.2.1, BL8.2.2, BL5.0.1 and BL5.0.3 beamlines at the Advanced Light Source in Berkeley, using crystals flash-frozen in the crystallization buffers supplemented with 15–20% glycerol at −170 °C.

Crystal structure determination. The TIR1–ASK1 structure was determined by the multiple-isomorphous-replacement method using native crystals and the five heavy-metal-compound-derived crystals mentioned above. All data sets were integrated and scaled using HKL2000[38]. A single multiple-isomorphous-replacement solution was found using the SOLVE program[39], with several heavy-metal sites located by the program for each derivative crystal. Using this phase information, the structural model was built manually in the program O[40] and refined using the program CNS[41]. Further refinement was done in REFMAC[42] with the TLS program parameters generated by the TLSMD server[43]. The structures of TIR1–ASK1 in complex with different combinations of auxins and AUX/IAA peptide were solved by molecular replacement with the TIR1–ASK1 structure using the Phaser program and refined using CNS and REFMAC.

Received 27 January; accepted 8 March 2007.

1. Woodward, A. W. & Bartel, B. Auxin: regulation, action, and interaction. *Ann. Bot. (Lond.)* **95,** 707–735 (2005).
2. Teale, W. D., Paponov, I. A. & Palme, K. Auxin in action: signalling, transport and the control of plant growth and development. *Nature Rev. Mol. Cell Biol.* **7,** 847–859 (2006).
3. Dharmasiri, N. & Estelle, M. Auxin signaling and regulated protein degradation. *Trends Plant Sci.* **9,** 302–308 (2004).
4. Gray, W. M., Kepinski, S., Rouse, D., Leyser, O. & Estelle, M. Auxin regulates SCF(TIR1)-dependent degradation of AUX/IAA proteins. *Nature* **414,** 271–276 (2001).
5. Hagen, G. & Guilfoyle, T. Auxin-responsive gene expression: genes, promoters and regulatory factors. *Plant Mol. Biol.* **49,** 373–385 (2002).
6. Reed, J. W. Roles and activities of Aux/IAA proteins in *Arabidopsis. Trends Plant Sci.* **6,** 420–425 (2001).
7. Liscum, E. & Reed, J. W. Genetics of Aux/IAA and ARF action in plant growth and development. *Plant Mol. Biol.* **49,** 387–400 (2002).
8. Zenser, N., Ellsmore, A., Leasure, C. & Callis, J. Auxin modulates the degradation rate of Aux/IAA proteins. *Proc. Natl Acad. Sci. USA* **98,** 11795–11800 (2001).
9. Tiwari, S. B., Wang, X. J., Hagen, G. & Guilfoyle, T. J. AUX/IAA proteins are active repressors, and their stability and activity are modulated by auxin. *Plant Cell* **13,** 2809–2822 (2001).
10. Remington, D. L., Vision, T. J., Guilfoyle, T. J. & Reed, J. W. Contrasting modes of diversification in the Aux/IAA and ARF gene families. *Plant Physiol.* **135,** 1738–1752 (2004).
11. Overvoorde, P. J. *et al.* Functional genomic analysis of the AUXIN/INDOLE-3-ACETIC ACID gene family members in *Arabidopsis* thaliana. *Plant Cell* **17,** 3282–3300 (2005).
12. Okushima, Y. *et al.* Functional genomic analysis of the AUXIN RESPONSE FACTOR gene family members in *Arabidopsis* thaliana: unique and overlapping functions of ARF7 and ARF19. *Plant Cell* **17,** 444–463 (2005).
13. Ramos, J. A., Zenser, N., Leyser, O. & Callis, J. Rapid degradation of auxin/indole acetic acid proteins requires conserved amino acids of domain II and is proteasome dependent. *Plant Cell* **13,** 2349–2360 (2001).
14. Dharmasiri, N., Dharmasiri, S., Jones, A. M. & Estelle, M. Auxin action in a cell-free system. *Curr. Biol.* **13,** 1418–1422 (2003).
15. Kepinski, S. & Leyser, O. Auxin-induced SCFTIR1-Aux/IAA interaction involves stable modification of the SCFTIR1 complex. *Proc. Natl Acad. Sci. USA* **101,** 12381–12386 (2004).
16. Dharmasiri, N., Dharmasiri, S. & Estelle, M. The F-box protein TIR1 is an auxin receptor. *Nature* **435,** 441–445 (2005).
17. Kepinski, S. & Leyser, O. The *Arabidopsis* F-box protein TIR1 is an auxin receptor. *Nature* **435,** 446–451 (2005).
18. Gagne, J. M., Downes, B. P., Shiu, S. H., Durski, A. M. & Vierstra, R. D. The F-box subunit of the SCF E3 complex is encoded by a diverse superfamily of genes in *Arabidopsis. Proc. Natl Acad. Sci. USA* **99,** 11519–11524 (2002).
19. Dharmasiri, N. *et al.* Plant development is regulated by a family of auxin receptor F box proteins. *Dev. Cell* **9,** 109–119 (2005).
20. Jonsson, A. *Encyclopaedia of Plant Physiology* **14,** 959–1006 (Springer, Berlin, 1961).
21. Kaethner, T. Conformational change theory for auxin structure-activity relationships. *Nature* **267,** 19–23 (1977).
22. Farrimond, J. A., Elliott, M. C. & Clack, D. W. Charge separation as a component of the structural requirements for hormone activity. *Nature* **274,** 401–402 (1978).
23. Petroski, M. D. & Deshaies, R. J. Function and regulation of cullin-RING ubiquitin ligases. *Nature Rev. Mol. Cell Biol.* **6,** 9–20 (2005).
24. Schulman, B. A. *et al.* Insights into SCF ubiquitin ligases from the structure of the Skp1-Skp2 complex. *Nature* **408,** 381–386 (2000).
25. Zheng, N. *et al.* Structure of the Cul1-Rbx1-Skp1-F boxSkp2 SCF ubiquitin ligase complex. *Nature* **416,** 703–709 (2002).
26. Kobe, B. & Kajava, A. V. The leucine-rich repeat as a protein recognition motif. *Curr. Opin. Struct. Biol.* **11,** 725–732 (2001).

27. Kobe, B. & Deisenhofer, J. Crystallization and preliminary X-ray analysis of porcine ribonuclease inhibitor, a protein with leucine-rich repeats. *J. Mol. Biol.* **231,** 137–140 (1993).
28. Choe, J., Kelker, M. S. & Wilson, I. A. Crystal structure of human toll-like receptor 3 (TLR3) ectodomain. *Science* **309,** 581–585 (2005).
29. Irvine, R. F. & Schell, M. J. Back in the water: the return of the inositol phosphates. *Nature Rev. Mol. Cell Biol.* **2,** 327–338 (2001).
30. Macbeth, M. R. *et al.* Inositol hexakisphosphate is bound in the ADAR2 core and required for RNA editing. *Science* **309,** 1534–1539 (2005).
31. Hanakahi, L. A., Bartlet-Jones, M., Chappell, C., Pappin, D. & West, S. C. Binding of inositol phosphate to DNA-PK and stimulation of double-strand break repair. *Cell* **102,** 721–729 (2000).
32. Steger, D. J., Haswell, E. S., Miller, A. L., Wente, S. R. & O'Shea, E. K. Regulation of chromatin remodeling by inositol polyphosphates. *Science* **299,** 114–116 (2003).
33. York, J. D., Odom, A. R., Murphy, R., Ives, E. B. & Wente, S. R. A phospholipase C-dependent inositol polyphosphate kinase pathway required for efficient messenger RNA export. *Science* **285,** 96–100 (1999).
34. Hoy, M. *et al.* Inositol hexakisphosphate promotes dynamin I-mediated endocytosis. *Proc. Natl Acad. Sci. USA* **99,** 6773–6777 (2002).
35. Moon, J., Parry, G. & Estelle, M. The ubiquitin-proteasome pathway and plant development. *Plant Cell* **16,** 3181–3195 (2004).
36. Xie, D. X., Feys, B. F., James, S., Nieto-Rostro, M. & Turner, J. G. COI1: an *Arabidopsis* gene required for jasmonate-regulated defense and fertility. *Science* **280,** 1091–1094 (1998).
37. Nalepa, G., Rolfe, M. & Harper, J. W. Drug discovery in the ubiquitin-proteasome system. *Nature Rev. Drug Discov.* **5,** 596–613 (2006).
38. Otwinowski, Z. & Minor, W. (eds) *Processing of X-ray Diffraction Data Collected in Oscillation Mode* (Academic Press, New York, 1997).
39. Terwilliger, T. C. Maximum-likelihood density modification. *Acta Crystallogr. D* **56,** 965–972 (2000).
40. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47,** 110–119 (1991).
41. Brunger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54,** 905–921 (1998).
42. CCP4. The CCP4 Suite: programs for protein crystallography. *Acta Crystallogr. D* **50,** 760–763 (1994).
43. Painter, J. & Merritt, E. A. Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr. D* **62,** 439–450 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the beamline staff of the Advanced Light Source at Berkeley for help with data collection. We also thank J. Callis, W. Xu and members of the Zheng laboratory for discussions and help. This work is supported by grants from the National Institutes of Health (M.E), the Pew Scholar Program (N.Z.), the National Science Foundation (M.E.) and the Department of Energy (M.E.).

Author Information Structure coordinates and structural factors are deposited in the Protein Data Bank under accession numbers 2P1M, 2P1N, 2P1O, 2P1P and 2P1Q (see Supplementary Table 1). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to N.Z. (nzheng@u.washington.edu).

Vol 446 | 5 April 2007 | doi:10.1038/nature05718

LETTERS

Global warming and climate forcing by recent albedo changes on Mars

Lori K. Fenton[1], Paul E. Geissler[3] & Robert M. Haberle[2]

For hundreds of years, scientists have tracked the changing appearance of Mars, first by hand drawings and later by photographs[1,2]. Because of this historical record, many classical albedo patterns have long been known to shift in appearance over time. Decadal variations of the martian surface albedo are generally attributed to removal and deposition of small amounts of relatively bright dust on the surface. Large swaths of the surface (up to 56 million km^2) have been observed to darken or brighten by 10 per cent or more[3–5]. It is unknown, however, how these albedo changes affect wind circulation, dust transport and the feedback between these processes and the martian climate. Here we present predictions from a Mars general circulation model, indicating that the observed interannual albedo alterations strongly influence the martian environment. Results indicate enhanced wind stress in recently darkened areas and decreased wind stress in brightened areas, producing a positive feedback system in which the albedo changes strengthen the winds that generate the changes. The simulations also predict a net annual global warming of surface air temperatures by ~0.65 K, enhancing dust lifting by increasing the likelihood of dust devil generation. The increase in global dust lifting by both wind stress and dust devils may affect the mechanisms that trigger large dust storm initiation, a poorly understood phenomenon, unique to Mars. In addition, predicted increases in summertime air temperatures at high southern latitudes would contribute to the rapid and steady scarp retreat that has been observed in the south polar residual ice for the past four Mars years[6–8]. Our results suggest that documented albedo changes affect recent climate change and large-scale weather patterns on Mars, and thus albedo variations are a necessary component of future atmospheric and climate studies.

Thermal Emission Spectrometer (TES) broadband albedos from the years 1999–2000 are shown in Fig. 1a, which has been artificially tinted a rust colour that mimics the apparent visible colour of Mars. In Fig. 1b, changes in albedo since the Viking mission are highlighted. Many areas on Mars darkened at some time during the 20-year interval between the Viking and Mars Global Surveyor (MGS) missions, particularly those regions associated with high thermal inertia[4]. Northern dark areas, such as Acidalia Planitia and northern Utopia

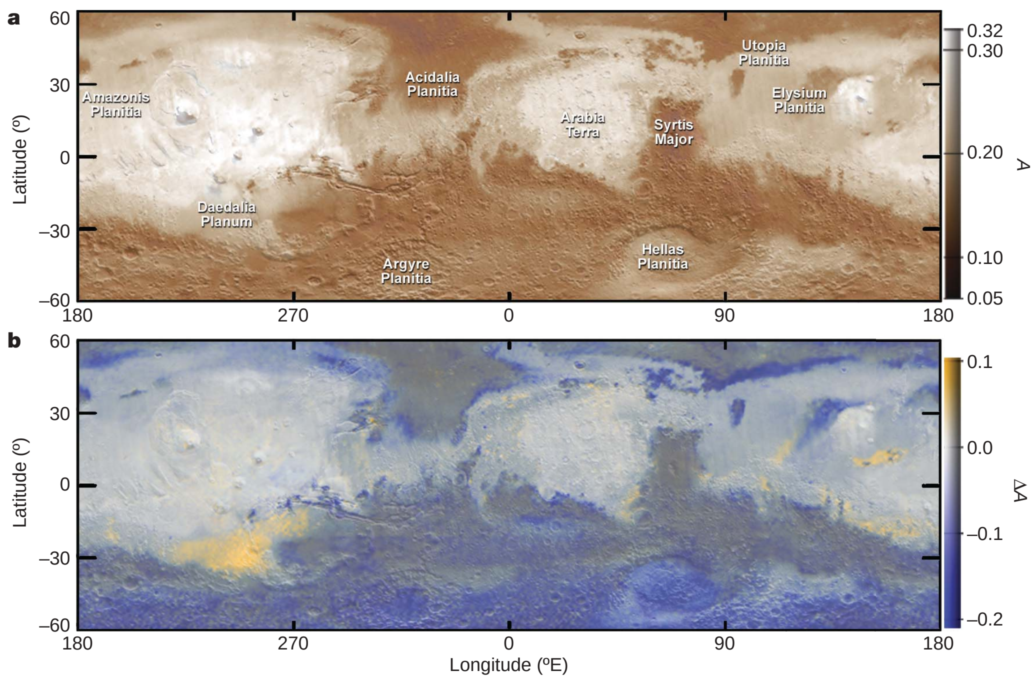


Figure 1 | Albedo and changes in albedo. **a**, Albedo map of Mars superimposed on Mars Orbiter Laser Altimeter (MOLA) shaded relief. Albedos (A) are calculated from the broadband TES bolometer: the rust tint is artificially applied to approximate the appearance of the martian surface. **b**, TES albedos in greyscale superimposed on MOLA shaded relief, with changes (ΔA) relative to the Viking IRTM shown in yellow for relative brightening and blue for relative darkening (see scale on right).

[1]Carl Sagan Center, [2]Space Science Division, NASA Ames Research Center, Moffett Field, California 94035, USA. [3]US Geological Survey, Flagstaff, Arizona 86001, USA.

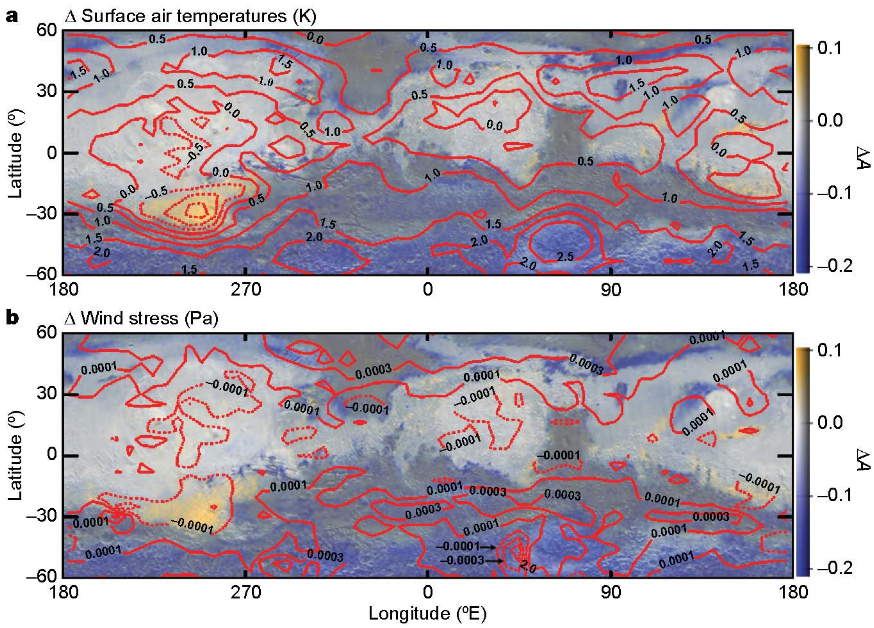


Figure 2 | The effect of albedo changes on air temperature and wind stress. **a**, MGCM-predicted annual-mean change in surface air temperatures (TES−IRTM; contours) on the same plot from Fig. 1. Note the temperature increases over darkened surfaces and temperature decreases over brightened surfaces. **b**, MGCM-predicted annual-mean change in wind stress (TES−IRTM; contours). Note the general correlation of wind stress increases over darkened surfaces and wind stress decreases over brightened surfaces.

Planitia, expanded southwards. Nearly the entire southern mid-latitude zone between $-40°$ and $-70°$ darkened, including Argyre and Hellas Planitiae. This darkening of the Southern Highlands is especially important because of the enhanced insolation during southern summer, when Mars is closest to the Sun. Most of the areas that brightened are located on the southern edges of high-albedo regions, such as Daedalia Planum. The only large regions that remained unchanged are the bright, dust mantled, low thermal inertia 'continents', such as Amazonis Planitia and Arabia Terra.

Earlier comparisons of atmospheric temperatures derived from MGS and Viking data indicated that most differences were caused by the temporally and spatially varying patterns of suspended atmospheric dust[9]. As the MGS mapping mission continued, further studies revealed differences not caused by atmospheric dust loading: for example, night-time atmospheric temperatures are repeatable from one Mars year to the next, but daytime atmospheric temperatures are more variable[10,11]. In particular, measurements from TES following the 2001 global dust storm (GDS) indicate that 4 K cooler daytime atmospheric temperatures persisted throughout the aphelion season following the storm, despite the facts that atmospheric dust opacities had long since dropped to pre-storm values and that night-time temperatures appeared consistent with previous years. This post-storm daytime cooling was attributed to deposition of bright dust on the surface after the dust storm, causing the albedo to increase, and subsequently reducing both surface and atmospheric temperatures[11]. Thus, interannual variations in dust fallout on the martian surface have had a measurable effect on atmospheric temperatures.

We applied two maps of surface Lambert albedo to the NASA Ames Mars general circulation model (MGCM)[12]; we used version 1.7.3 of the MGCM with an Arakawa C-grid (in which zonal and meridional winds are staggered horizontally between the main longitudinal and latitudinal grid point positions, respectively). One map was produced from the Infrared Thermal Mapper (IRTM) on the Viking Orbiters, derived from calibrated data obtained during non-dusty periods of the first and second years of the Viking mission (1976–78)[3,13,14]. The second map was produced from TES albedos derived from calibrated data obtained during non-dusty periods during the first year of the MGS mapping mission (1999–2000)[15]. The two measurements span very similar wavelength ranges (0.3–3.0 μm for IRTM[16] and 0.3–2.9 μm for TES[15]), and thus are directly comparable. Because the original IRTM albedo map spans latitudes between $\pm 60°$, TES albedos were substituted in areas poleward of this boundary. Thus, there is no effective change in albedo in the polar regions and higher latitudes in these model runs. In each case, the MGCM was run for a full model year (following a spin-up year), with a horizontal grid spacing of 6° of longitude by 5° of latitude, and 24 vertical atmospheric levels ranging from the surface to a pressure of 0.0005 mbar ($\sim$100 km). The visible dust opacity was set to a constant value of 0.3 on a 6.1-mbar pressure surface. Although a simplification of the martian dust cycle, this dust scheme ensures that differences in wind circulation and other parameters are caused solely by surface forcing from albedo changes.

Differencing of atmospheric parameters from the MGCM output demonstrates that the albedo fluctuations do drive some atmospheric changes. Table 1 lists annual global-mean changes of several parameters, calculated by averaging the difference of each parameter in each grid point and time step, weighted by the zonal-mean spatial area in each grid point. Although the changes are slight, annual

Table 1 | Annual global-mean values predicted by the MGCM

Parameter	IRTM	TES	$\Delta_{TES-IRTM}$	$\%\Delta_{TES-IRTM}$
Ground temperature	201.58 K	202.43 K	+0.86 K	+0.42%
Surface air temperature	195.61 K	196.26 K	+0.65 K	+0.33%
Condensed CO_2	5.83×10^{12} kg	5.79×10^{12} kg	-3.62×10^{10} kg	−0.62%
Daytime planetary boundary layer	0.85 km	0.88 km	+0.02 km	+2.7%
Wind stress	3.49×10^{-3} Pa	3.56×10^{-3} Pa	$+0.07 \times 10^{-3}$ Pa	+2.9%
Surface wind speed	6.76 m s^{-1}	6.84 m s^{-1}	+0.08 m s^{-1}	+1.3%

global-mean values are not representative of more dramatic changes on smaller spatial and temporal scales. Furthermore, small but steady changes acting on a timescale of decades could lead to much larger cumulative shifts in climate-influencing processes. Ground and surface air temperatures increase slightly as a result of the global-mean decrease in albedo and subsequent increase in absorbed solar energy on the surface. In particular, the surface air temperature increase of ~0.65 K is a value on the same scale as that of measured interannual variations of land-surface air temperatures on Earth, even though the processes involved are significantly different[17]. Increased heating near the surface leads to greater convective instability, pushing the planetary boundary layer slightly higher. Surface wind stresses increase slightly, largely as a result of increases in the magnitude of near-surface wind velocity. Because the observed global annual-mean differences in Table 1 are small, sensitivity tests of the MGCM are useful in determining the reliability of these numbers. Adding an initial perturbation or running at a different resolution produced changes that were at least an order of magnitude smaller than those produced by albedo changes, and so the model results presented here are considered robust.

Figure 2 shows the albedo-driven differences in annual-mean surface atmospheric temperatures and wind stresses as a function of latitude and longitude. Warmer temperatures and higher wind stresses correlate with the areas where the albedo decreased, and cooler temperatures and lower wind stresses generally correlate with areas that brightened. Previous work has shown that martian albedo changes within the range discussed in this work can locally change the predicted planetary boundary layer depth by more than a kilometre, which could significantly influence dust devil generation[18]. One model of dust devil formation treats these vortices as heat engines, with a thermodynamic efficiency related to the depth of the planetary boundary layer[19]. Thus a decrease in surface albedo could contribute to enhanced dust lifting for two reasons: (1) increased wind stresses directly lift more particulate material (both dust and sand) off the surface, and (2) a more unstable atmosphere will lead to a higher dust devil generation rate, lifting more dust into the atmosphere. The coupling of these processes with albedo changes could produce a positive feedback mechanism of surface–atmosphere interaction, in which surface albedo reductions enhance the dust lifting processes that produce the surface changes.

Moreover, a change in global surface albedo could influence the formation of dust storms, both on local and global scales. It has been suggested that the surface brightening and atmospheric cooling following the 2001 GDS may affect the timing of future large dust storms[11]. Indeed, our work shows that a surface brightening caused by dust fallout would decrease wind stresses and suppress dust devil formation, the two mechanisms potentially responsible for dust storm initiation. Our results indicate that widespread dust fallout from one GDS-bearing year would make it more difficult for a GDS to occur in the following years. The lack of continuity in the historical record of GDS occurrence makes it difficult to test this hypothesis, but the exceptions of large storms in four successive Mars years from 1971–77[20,21] indicate that other factors must also be in play and able to overcome (on occasion in a single year) the significant effects of albedo changes due to dust deposition. The relatively high number of dust storms of the 1970s appears to have contributed to higher Viking albedos relative to those measured by MGS in 1999–2000, which followed a decade or more of relative inactivity[5]. If global surface albedo affects dust storm initiation (and is itself affected by dust storm fallout), then long term trends in albedo change could dictate decadal periods prone to large dust storms versus periods free of such storms. Although the events that trigger dust storms (in particular, GDSs) have yet to be fully understood, this work demonstrates that one contributing factor may be a decrease in surface albedo. Thus, a realistically time-varying surface albedo must be included in future atmospheric simulations in order to accurately represent the dust cycle.

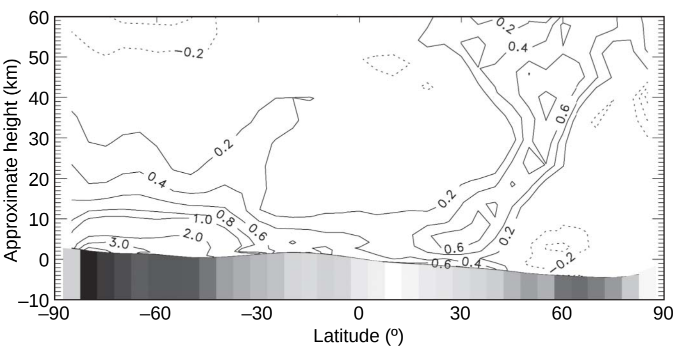


Figure 3 | MGCM air temperature differences during perihelion season. Zonal-mean air temperature differences (TES−IRTM) as a function of approximate elevation and latitude for $L_s = 260°–280°$. Raw temperatures were interpolated onto constant pressure surfaces before averaging. Zonal-mean surface level is plotted at the bottom, with zonal-mean differences in albedo changes shown in greyscale beneath the surface (white, $\Delta A = +0.001$; black, $\Delta A = -0.088$; MOLA elevation datum refers to ~6.1-mbar pressure level). Note that contours smaller than 1 K are shown in 0.2 K intervals; contours greater than 1 K are shown in 1 K intervals. Air temperatures above the south polar residual cap increase by 3–4 K near the surface, probably enhancing scarp retreat in the residual polar ice cap.

Erosion of depression walls in the southern residual polar cap has been observed by the Mars Orbiter Camera (MOC) during the MGS mission, suggesting that short-term climate change is currently occurring on Mars[6]. Further study has led to the proposal that over the last 100–150 martian years the southern residual polar cap has undergone a series of short depositional periods followed by longer periods (on the order of one or more decades) of erosion by scarp retreat[8]. According to this analysis, the most recent CO_2 ice layer may have been deposited shortly after 1972. It is postulated that these periods of deposition and erosion may be linked to dust storm frequency or changes in surface albedo[8]. If dust storms were responsible, then the youngest residual ice layer would have been deposited as a result of atmospheric changes caused by the 1971 GDS. However, the relationship between GDSs and polar cap erosion is problematic, because multiple year coverage from MOC indicates that no permanent ice layer was deposited in the aphelion season (southern winter) after the 2001 GDS[6], despite the storm-related atmospheric cooling and surface brightening[10,11]. It is also likely that no additions were made to the residual ice cap in winters following the 1973–77 GDSs. It appears that the southern residual ice cap was in a state of erosion during the period between the late 1970s up to the present.

To investigate albedo-driven changes at high latitudes, we reran the IRTM albedo case, this time including IRTM albedos spanning latitudes ±80°. IRTM albedos poleward of ±60° were derived using an atmospherically corrected subsurface conduction model[22–24]. Recently reported emission angle effects may cause the actual high-latitude TES–IRTM albedo differences to be smaller than those applied here[5], potentially lessening the following results. Figure 3 shows that zonal-mean summertime temperatures over the southern polar cap increased by 0.2–2 K poleward of −85° from the surface up to an altitude of ~30 km. The subsequent increase in emitted downwelling infrared radiation predicted by the MGCM in grid points spanning the residual polar cap is ~1 W m^{-2}. If this entire increase in infrared radiation goes into sublimating residual CO_2 ice, the potential mass loss during the southern summer (when the solar longitude $L_s = 250°–350°$) is $\sim 5 \times 10^{12}$ kg. Previous estimates of summertime CO_2 mass loss from the south polar residual cap, based on observations of scarp retreat, range from 8.4×10^{12} kg (ref. 8) to $(2–4) \times 10^{13}$ kg (ref. 6). Assuming that all other polar processes (for example, cloud condensation, spatially varying atmospheric dust) cancel out between the IRTM and TES model runs, surface darkening at high southern latitudes could contribute substantially

to permanent erosion of CO_2 deposits, probably preventing recondensation in any potential cold traps.

Martian climate indicators, such as GDS occurrence, polar energy balance, and annual global-mean air temperature, are dependent on many interrelated and poorly understood processes. By investigating solely the effects of changes in surface albedo (from two very different Mars years), we have shown that albedo interacts with, and could in part drive, other climate-influencing processes on Mars.

Received 6 September 2006; accepted 22 February 2007.

1. Flammarion, C. *La Planète Mars et ses Conditions d'Habitabilité* Vols 1 and 2 (Gauthier Villars et Fils, Paris, 1892).
2. de Vaucouleurs, G. *Physics of the Planet Mars* (Faber and Faber, London, 1954).
3. Christensen, P. R. Global albedo variations on Mars: Implications for active aeolian transport, deposition, and erosion. *J. Geophys. Res.* **93** (B7), 7611–7624 (1988).
4. Geissler, P. E. Three decades of Martian surface changes. *J. Geophys. Res.* **110,** E02001, doi:10.1029/2004JE002345 (2005).
5. Szwast, M. A., Richardson, M. I. & Vasavada, A. R. Surface dust redistribution on Mars as observed by the Mars Global Surveyor and Viking Orbiters. *J. Geophys. Res.* **111,** E11008, doi:10.1029/2005JE002485 (2006).
6. Malin, M. C., Caplinger, M. A. & Davis, S. D. Observational evidence for an active surface reservoir of solid carbon dioxide on Mars. *Science* **294,** 2146–2148 (2001).
7. Byrne, S. & Ingersoll, A. P. A sublimation model for Martian south polar ice features. *Science* **299,** 1051–1053 (2003).
8. Thomas, P. C. *et al.* South polar residual cap of Mars: Features, stratigraphy, and changes. *Icarus* **174,** 535–559 (2005).
9. Clancy, R. T. *et al.* An intercomparison of ground-based millimetre, MGS TES, and Viking atmospheric temperature measurements: Seasonal and interannual variability of temperatures and dust loading in the global Mars atmosphere. *J. Geophys. Res.* **105** (E4), 9553–9571 (2000).
10. Liu, J., Richardson, M. I. & Wilson, R. J. An assessment of the global, seasonal, and interannual spacecraft record of Martian climate in the thermal infrared. *J. Geophys. Res.* **108** (E8), 5089, doi:10.1029/2002JE001921 (2003).
11. Smith, M. D. Interannual variability in TES atmospheric observations of Mars during 1999–2003. *Icarus* **167,** 148–165 (2004).
12. Haberle, R. M. *et al.* General circulation model simulations of the Mars Pathfinder atmospheric structure investigation/meteorology data. *J. Geophys. Res.* **104** (E4), 8957–8974 (1999).
13. Kieffer, H. H. *et al.* Thermal and albedo mapping of Mars during the Viking primary mission. *J. Geophys. Res.* **82,** 4249–4292 (1977).
14. Pleskot, L. K. & Miner, E. D. Time variability of martian bolometric albedo. *Icarus* **45,** 179–201 (1981).
15. Christensen, P. R. *et al.* Mars Global Surveyor Thermal Emission Spectrometer experiment: Investigation description and surface science results. *J. Geophys. Res.* **106** (E10), 23823–23872 (2001).
16. Kieffer, H. H., Davis, P. A. & Soderblom, L. A. Mars' global properties — Maps and applications. *Proc. Lunar Planet. Sci. Conf.* **XII,** 1395–1417 (1982).
17. Houghton, J. T., *et al. Climate Change: The Scientific Basis* 105–107 (Cambridge Univ. Press, New York, 2001).
18. Kahre, M. A. *et al.* Observing the martian surface albedo pattern: Comparing the AEOS and TES data sets. *Icarus* **179,** 55–62 (2005).
19. Rennó, N. O., Burkett, M. L. & Larkin, M. P. A simple thermodynamical theory for dust devils. *J. Atmos. Sci.* **55,** 3244–3252 (1998).
20. Zurek, R. W. & Martin, L. J. Interannual variability of planet-encircling dust storms on Mars. *J. Geophys. Res.* **98** (E2), 3247–3259 (1993).
21. McKim, R. Telescopic Martian dust storms: A narrative and catalogue. *Mem. Br. Astron. Assoc.* **44,** 1–168 (1999).
22. Vasavada, A. R. *et al.* Surface properties of Mars' polar layered deposits and polar landing sites. *J. Geophys. Res.* **105** (E3), 6961–6970 (1999).
23. Paige, D. A., Bachman, J. E. & Keegan, K. D. Thermal and albedo mapping of the polar regions of Mars using Viking thermal mapper observations, 1, North polar region. *J. Geophys. Res.* **99** (E12), 25959–25992 (1994).
24. Paige, D. A. & Keegan, K. D. Thermal and albedo mapping of the polar regions of Mars using Viking thermal mapper observations, 2, South polar region. *J. Geophys. Res.* **99** (E12), 25993–26014 (1994).

Acknowledgements We thank J. Schaeffer for help with the MGCM, and T. Michaels and R. Zurek for comments and suggestions. This work was supported by the NASA Mars Data Analysis Program.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to L.K.F. (fenton@mintz.arc.nasa.gov) or P.E.G. (pgeissler@usgs.gov).

LETTERS

Lanthanide contraction and magnetism in the heavy rare earth elements

I. D. Hughes[1], M. Däne[2], A. Ernst[3], W. Hergert[2], M. Lüders[4], J. Poulter[5], J. B. Staunton[1], A. Svane[6], Z. Szotek[4] & W. M. Temmerman[4]

The heavy rare earth elements crystallize into hexagonally close packed (h.c.p.) structures and share a common outer electronic configuration, differing only in the number of 4*f* electrons they have[1]. These chemically inert 4*f* electrons set up localized magnetic moments, which are coupled via an indirect exchange interaction involving the conduction electrons. This leads to the formation of a wide variety of magnetic structures, the periodicities of which are often incommensurate with the underlying crystal lattice[2]. Such incommensurate ordering is associated with a 'webbed' topology[3,4] of the momentum space surface separating the occupied and unoccupied electron states (the Fermi surface). The shape of this surface—and hence the magnetic structure—for the heavy rare earth elements is known to depend on the ratio of the interplanar spacing *c* and the interatomic, intraplanar spacing *a* of the h.c.p. lattice[5]. A theoretical understanding of this problem is, however, far from complete. Here, using gadolinium as a prototype for all the heavy rare earth elements, we generate a unified magnetic phase diagram, which unequivocally links the magnetic structures of the heavy rare earths to their lattice parameters. In addition to verifying the importance of the *c*/*a* ratio, we find that the atomic unit cell volume plays a separate, distinct role in determining the magnetic properties: we show that the trend from ferromagnetism to incommensurate ordering as atomic number increases is connected to the concomitant decrease in unit cell volume. This volume decrease occurs because of the so-called lanthanide contraction[6], where the addition of electrons to the poorly shielding 4*f* orbitals leads to an increase in effective nuclear charge and, correspondingly, a decrease in ionic radii.

We report here a theoretical investigation of the onset of magnetic order in the heavy rare earths, focusing on gadolinium as the prototypical member of this series. State-of-the-art computational techniques are used to model the finite-temperature electronic structure of this system, with a completely *ab initio* approach taken and no fitting to experimental parameters. Thermally induced spin fluctuations are treated using a 'local moment' picture of magnetism[7], in which local spin polarization axes are associated with all lattice sites, the orientations of which vary slowly on the timescale of electronic motion. These 'local moment' degrees of freedom produce local magnetic fields on the lattice sites, which affect the electronic motions and are self-consistently maintained by them. By studying the wave vector (**q**) dependence of the spin fluctuations that characterize the paramagnetic state we gain information about the type of magnetic order that might occur as the temperature is lowered through a phase transition.

For example, in a ferromagnetic material such as gadolinium, the paramagnetic state is characterized by ferromagnetic spin fluctuations that have long wavelengths ($\mathbf{q} \approx 0$) and becomes unstable to these fluctuations at the Curie temperature T_C. For a system that orders into an incommensurate antiferromagnetic structure the paramagnetic state is dominated by 'anti-ferromagnetic' spin fluctuations, specified by a finite, incommensurate, wave vector $\mathbf{q} = \mathbf{Q}_0$ that also characterizes the static magnetization or spin density wave state formed below the Néel temperature T_N. For example, the magnetic structures of the later heavy rare earth elements, terbium to thulium, are described by wave vectors of the form $\mathbf{Q}_0 = \{0, 0, q_{inc}\}$, where individual hexagonal layers are uniformly magnetized in a direction that changes from layer to layer according to the modulation vector q_{inc}. To determine the nature of the spin fluctuations we evaluate the paramagnetic spin susceptibility, $\chi(\mathbf{q}, T)$. This can be written as $\chi(\mathbf{q}, T) = \mu^2/(3k_BT - S^{(2)}(\mathbf{q}, T))$, where μ specifies a local moment magnitude and $S^{(2)}(\mathbf{q}, T)$ mediates the interaction between moments[8]. The wave vector at which $\chi(\mathbf{q}, T)$ and $S^{(2)}(\mathbf{q}, T)$ attain their maxima corresponds to the wave vector of the dominant paramagnetic spin fluctuations.

We started by investigating gadolinium at its equilibrium lattice parameters, both those measured experimentally and also those determined theoretically by minimizing the total energy of the system with respect to changes in the parameters. For a proper description of the electronic structure of gadolinium it was necessary to account for the strong electron–electron correlations of the highly localized 4*f* states. Indeed, on neglecting these strong electron correlations the *f*-electrons became band-like and we found that at both the experimental and theoretical lattice parameters the system was inclined to order into a commensurate antiferromagnetic structure, with magnetic moments oppositely aligned in alternating planes along the *c* axis, contrary to the experimentally observed ferromagnetic ordering. Such antiferromagnetic ordering has been found in other conventional electronic structure calculations[9] and is attributed to the presence of minority spin 4*f* electrons at the Fermi energy[10]. To localize the *f* electrons, previous investigations separated them out from the more itinerant *spd* electrons, treating them either as part of the core[11] or introducing the effect of strong electron correlations by explicitly including a Coulomb parameter *U* for the *f* states[12]. In our investigation we used a different approach, the self-interaction correction technique[13–15] (see Supplementary Methods for details), in which localized and delocalized electrons are treated on an equal footing and the effects of strong Coulomb correlations are incorporated in a parameter-free way.

The picture that emerged from our first-principles theory for $S^{(2)}(\mathbf{q}, T)$, once the *f* electrons were appropriately localized, was of a magnetic field produced by a 4*f* moment on one atomic site polarizing the conduction electrons. Their induced magnetization then

[1]Department of Physics, University of Warwick, Gibbet Hill Road, Coventry, CV4 7AL, UK. [2]Institut für Physik, Martin-Luther-Universität Halle-Wittenberg, Friedemann-Bach-Platz 6, D-06099 Halle, Germany. [3]Max Planck Institut für Mikrostrukturphysik, Weinberg 2, D-06120 Halle, Germany. [4]Daresbury Laboratory, Daresbury, Warrington, WA4 4AD, UK. [5]Department of Mathematics, Faculty of Science, Mahidol University, Bangkok, 10400, Thailand. [6]Department of Physics and Astronomy, University of Aarhus, DK-8000 Aarhus, Denmark.

interacted with the magnetic fields set up by moments on other sites. Consequently, the moment–moment interaction was governed by the non-interacting susceptibility of the conduction electrons, $\chi_0^{\mathrm{conduction}}(\mathbf{q}, T)$. Such an indirect exchange interaction between moments, mediated by conduction electrons, is referred to as a RKKY (Ruderman–Kittel-Kasuya–Yosida) interaction and experimentally is what drives the magnetism in the heavy rare earth elements[16]. Indeed, with a proper treatment of the f electrons, our calculations predicted a ferromagnetic ground state, in agreement with experiment. The temperature dependency of the susceptibility followed a Curie–Weiss law, with $T_C = 280$ K for the theoretical lattice parameters and 324 K for the experimental lattice parameters. The overestimate of the Curie temperature at the experimental lattice parameters (experimental $T_C = 293$ K; ref. 2) can be attributed to our mean field treatment of spin fluctuations[17]. The effective magnetic moment was $7.34\mu_B$ for the theoretical lattice parameters and $7.36\mu_B$ for the experimental lattice parameters, in reasonable agreement with the experimental value of $7.63\mu_B$ and also the results of previous studies in which f electrons were treated in a different subsystem to the more itinerant electrons ($7.44\mu_B$ and $7.41\mu_B$; refs 18 and 10, respectively).

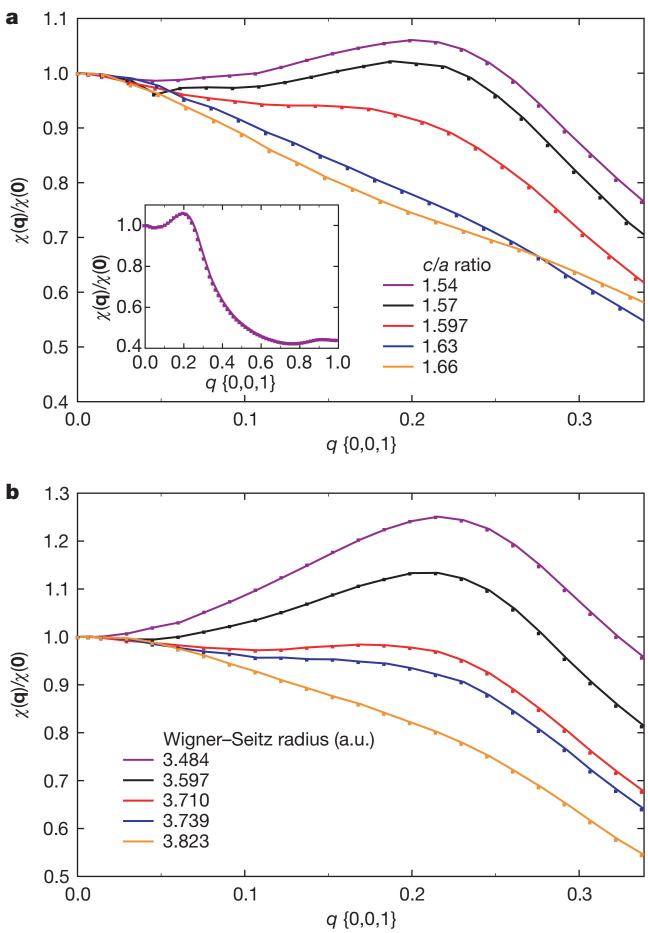


Figure 1 | Normalized paramagnetic spin susceptibilities for gadolinium, obtained from *ab initio* calculations. The data shown is for wave vectors $\mathbf{q}$ along the [0, 0, 1] direction. **a**, Data for various c/a ratios, using theoretical unit cell volumes which were found to be almost invariant to the c/a ratio. The inset shows the susceptibility up to the zone boundary for c/a ratio 1.54. The experimental c/a ratio is 1.597. It is evident that as the c/a ratio decreases, the susceptibility starts to develop a peak at some incommensurate wave vector q_{inc} and, at the lowest c/a ratios, the susceptibility no longer attains its maximum value at $\mathbf{q} = \mathbf{0}$. This means that, rather than ordering ferromagnetically, the system will adopt an incommensurate antiferromagnetic structure, modulated along the c axis with the wave vector q_{inc}. **b**, Data for various unit cell volumes, using a c/a ratio of 1.54. The unit cell volumes are parameterized using Wigner–Seitz (W–S) radii, defined as the radius of a sphere of the same volume as the volume per atom. Atomic units (a.u.) are used, where 1 a.u. is the Bohr radius of a hydrogen atom. As the volume is increased, the height of the incommensurate peak relative to the $\mathbf{q} = \mathbf{0}$ (ferromagnetic) peak is reduced and at a W–S radius of 3.710 a.u. there is a near degeneracy between the two ordering types. For the highest W–S radii the ferromagnetic peak wins out. A similiar trend was observed for other, small, c/a ratios that correspond to systems having a webbing feature in their Fermi surface.

The susceptibility of gadolinium for non-equilibrium lattice parameters is shown in Fig. 1. Two variables, the c/a ratio and the unit cell volume, specify the h.c.p. lattice parameters of the heavy rare earth elements. Figure 1a concerns the first. It shows that, for the lowest c/a ratios, gadolinium is no longer predicted to be ferromagnetic, instead adopting an incommensurate magnetic structure. This could be helical, where the helix turn angle, that is, the angle between magnetic moments in adjacent layers, would be given by πq_{inc}, where q_{inc} is the position of the susceptibility peak. Andrianov[19] investigated helical magnetic ordering in several heavy rare earth systems and found the helix turn angle to be a smooth, square-root-shaped function of the c/a ratio. He also noted that it varied by several orders of magnitude while the c/a ratio changed by less than 1%. Such behaviour could possibly be interpreted in terms of an electronic topological transition[20,21] at some critical c/a ratio, where this is a rupturing of the webbing structure found in the Fermi surface of these systems. This webbing structure contains large parallel sheets of Fermi surface, which can 'nest' together when translated by some vector in $\mathbf{k}$-space, and it has been shown, both theoretically[22] and experimentally[23], that the size of this nesting vector is correlated with the magnetic modulation vector.

In our theory, such Fermi surface effects should be contained in $\chi_0^{\mathrm{conduction}}(\mathbf{q}, T)$, because this is determined by excitations from occupied states at all Bloch wave vectors $\mathbf{k}$ to unoccupied states at wave vectors $\mathbf{k} + \mathbf{q}$. $\chi_0^{\mathrm{conduction}}(\mathbf{q}, T)$ is dominated by a Brillouin zone integral[17] $\int A_B(\mathbf{k}, \varepsilon_F) A_B(\mathbf{k} + \mathbf{q}, \varepsilon_F)\, d\mathbf{k}$ where the Bloch spectral function, $A_B(\mathbf{k}, \varepsilon_F)$, gives the density of states at the Fermi energy ε_F at the Bloch wave vector $\mathbf{k}$. It follows that it can develop a peak at a finite Bloch wave vector $\mathbf{q} = \mathbf{Q}_0$ if there are regions of the Fermi surface that coincide when one is translated by a nesting vector $\mathbf{Q}_0$. To analyse the results we obtained in Fig. 1a, we calculated the Fermi surface of paramagnetic gadolinium at various c/a ratios. A 'webbing' structure was indeed found in those systems where incommensurate ordering is predicted (see Supplementary Fig. S1 for three-dimensional illustrations of the Fermi surface), with the nesting vector magnitude coinciding with q_{inc}.

Figure 2 shows cross-sections through the nesting region of the Fermi surface for various c/a ratios. Two extremal nesting vectors, one centred and the other non-centred, were found. The length of the centred nesting vector decreases as the c/a ratio increases, in agreement with the experimental results of Andrianov[19], but contrary to the results in Fig. 1a, where the position of the incommensurate ordering peak is almost invariant to the c/a ratio used. The non-centred vector, however, stays fairly constant as the c/a ratio is altered, indicating this vector to be responsible for the incommensurate ordering observed in our calculations. This agrees with recent theoretical work[24].

We now turn to the magnetic ordering behaviour as a function of unit cell volume. Here we found two distinct cases, dependent on the c/a ratio of the lattice parameters. For high c/a ratios, corresponding to systems with no webbing feature, ferromagnetic ordering was predicted for all volumes. For low c/a ratios, corresponding to systems with webbing, a more complicated picture emerges, as shown in Fig. 1b. The webbing structure leads to an enhancement of the susceptibility at the nesting vector for all volumes. However, for the enhancement to be large enough so that incommensurate ordering wins out over ferromagnetic ordering the unit cell volume needs to be below a certain critical value.

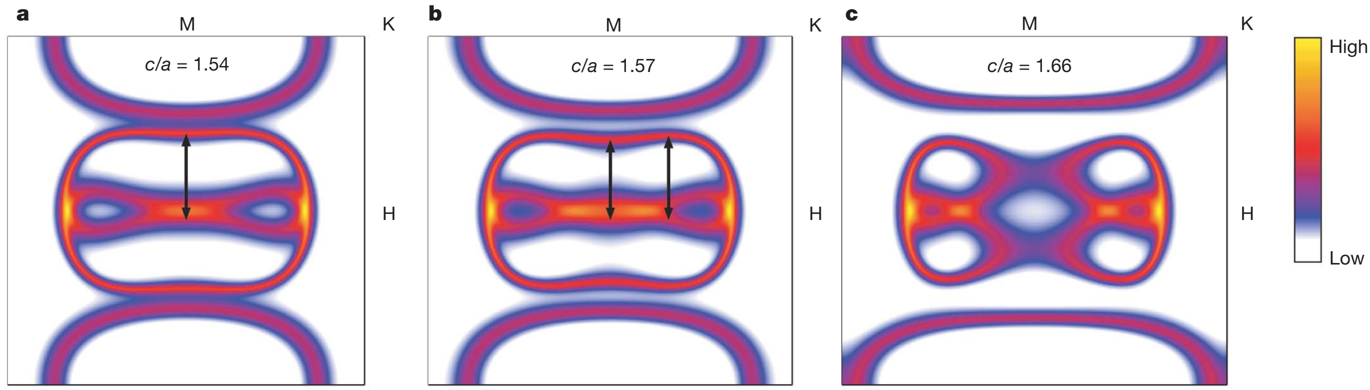


Figure 2 | Bloch spectral function of gadolinium on the HLMK plane of the hexagonal Brillouin zone, depicting the topology of the Fermi surface. The colour scale shows the spectral function in arbitrary units. Theoretical unit cell volumes were used, with the c/a ratios given in panels **a**, **b** and **c**. The centre of the plane is the L point. The Fermi surface in **a** has the 'webbing feature', with large sections of parallel Fermi surface that are separated by a nesting vector, indicated by the black arrow. The magnitude of the nesting vector is 0.2, which coincides with the size of the magnetic ordering wave vector $q_{\rm inc}$ that we obtained for this c/a ratio in Fig. 1a. As the c/a ratio increases, it is evident that the surfaces parallel to the KMK direction become less flat and at the highest c/a ratio the webbing structure is ruptured. This is concurrent with the transition to ferromagnetism observed in Fig. 1a.

Our key Fig. 3 summarizes the magnetic ordering tendencies of gadolinium as a function of c/a ratio and unit cell volume. Because 4f electrons are usually chemically inert, the heavy rare earth elements can all be considered as variations of the same entity and hence the behaviour of gadolinium as a function of lattice parameters is expected to mimic that of all the other heavy rare earth elements. Indeed, by examining where the experimental lattice parameters of all the heavy rare earth elements lie on Fig. 3, we predict that when going left to right in the heavy rare earth series there should be a trend away from ferromagnetism and towards incommensurate ordering. This is exactly what is observed experimentally, with the magnetic modulation vector starting out at zero for gadolinium (ferromagnetic ordering) and then progressively increasing through the series to produce various incommensurate antiferromagnetic structures.

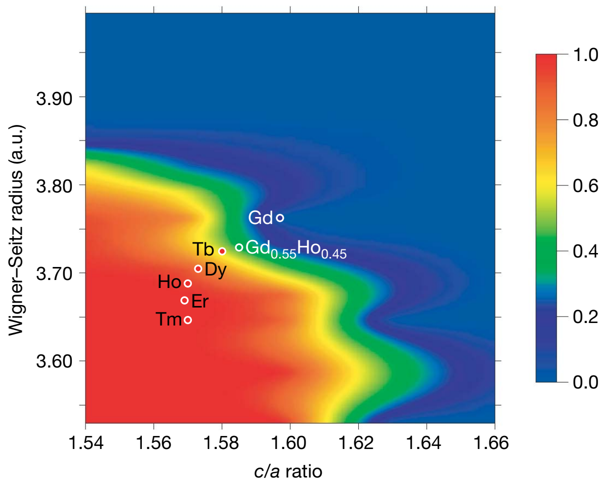


Figure 3 | Magnetic ordering tendencies of gadolinium (Gd) as a function of c/a ratio and W–S radii. On the colour scale, a value of 0 corresponds to when the paramagnetic state is dominated by ferromagnetic spin fluctuations, and a value of 1 corresponds to when it is dominated by spin fluctuations with some finite, incommensurate wave vector. For values between 0 and 1 the two ordering types compete: values below 0.5 indicate a stronger tendency towards ferromagnetic ordering and values above 0.5 indicate a greater tendency towards incommensurate antiferromagnetic ordering. A 3% shift was applied to the W–S radii, such that data shown at the experimental W–S radius of gadolinium corresponds to data calculated at the theoretical W–S radius of gadolinium. The experimental lattice parameters of all the heavy rare earth elements are indicated by circles: a blue (or red) circle indicates that experimentally the magnetic structure of the element is ferromagnetic (or incommensurate antiferromagnetic). The green circle indicates the experimental lattice parameters of a Gd–Ho alloy at the critical concentration of Ho at which an incommensurate antiferromagnetic phase first appears. From the phase diagram we predict the critical concentration of the $Gd_{1-x}Re_x$ alloys to be 0.78, 0.56, 0.49, 0.45, 0.42 for the heavy rare earth elements Tb (terbium), Dy (dysprosium), Ho (holmium), Er (erbium) and Tm (thulium) respectively. Experimental values are known only for dysprosium and holmium and are 0.50 (ref. 26) and 0.45 (ref. 27) respectively.

From our ordering phase diagram (Fig. 3), we predict that the transition between ferromagnetism and incommensurate ordering occurs very rapidly as a function of c/a ratio, particularly for the higher unit cell volumes. This is consistent with recent experimental work on terbium, for which it was shown[25] that its incommensurately ordered phase could be completely suppressed by increasing the c/a ratio by as little as 0.002. In the phase diagram the elements dysprosium and terbium are positioned close to, or within, the transition region between ferromagnetic and incommensurate ordering. This concurs with the two systems' experimental behaviour, which exhibits incommensurate ordering at high temperatures and ferromagnetic ordering at low temperatures. Gadolinium is able to form alloys with all the heavy rare earth elements, with the alloys transforming from ferromagnets to incommensurate magnetically structured materials once the concentration of the heavy rare earth element exceeds a critical value. We used the phase diagram to predict these critical alloy concentrations and found them to be in good agreement with experimental values where known (see Fig. 3 legend).

We also computed estimates of the magnetic ordering vectors of all the heavy rare earth elements from our susceptibility calculations for gadolinium, the results of which are shown in Fig. 4, where the experimental trend is well reproduced. The magnetic ordering vectors of the last three members of the series (holmium, erbium and thulium) were found to lie very close together, in agreement with experiment. Owing to its half-filled 4f shell, the gadolinium ion has orbital angular momentum $L = 0$. Hence the effects of spin orbit coupling can be neglected in our calculations of gadolinium. For the other heavy rare earth elements, however, the coupling of spin and orbital moments is important in obtaining estimates of their magnetic moments and magnetic ordering temperatures. Indeed, by accounting for this aspect, we used the gadolinium results to reproduce trends in magnetic ordering temperatures, as shown in the inset of Fig. 4. Nonetheless, the type of magnetic order and magnetic ordering vector are determined by the spd conduction electrons, which are little affected by spin-orbit coupling and which all the heavy rare earths have in common.

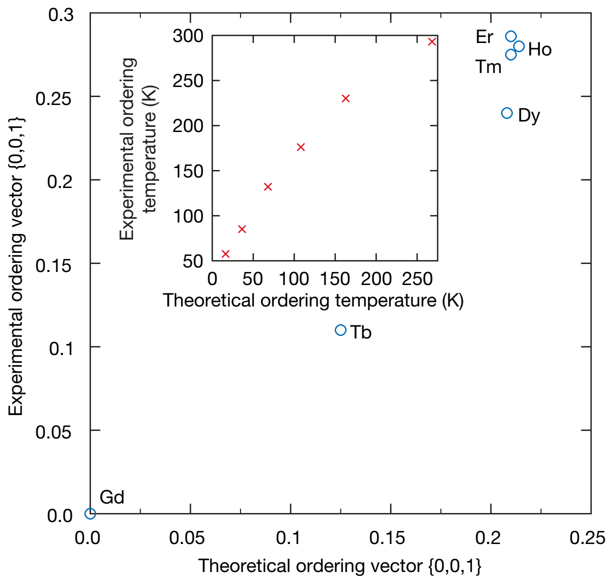


Figure 4 | Experimental magnetic ordering vectors of the heavy rare earth elements versus those predicted from *ab initio* calculations for gadolinium. The ordering vectors are obtained by performing susceptibility calculations for gadolinium at the appropriate lattice parameters. For example, when we performed a calculation at the experimental lattice parameters of terbium the susceptibility peaked at a wave vector $\mathbf{q} = \{0, 0, 0.13\}$, in good agreement with the experimental ordering vector, $\{0, 0, 0.11\}$. The inset shows the corresponding ordering temperatures. Experimentally, gadolinium has the highest ordering temperature, which decreases monotonically through the heavy rare earth series. To take into account the different total angular momentum values J of the heavy rare earth elements, the theoretical ordering temperatures have been scaled according to the de Gennes factor[28] $(g_J - 1)^2 J(J+1)$, where g_J is the Lande g-factor. It is clear that the results from the gadolinium study reproduce the experimental trend, although the magnitudes of the temperatures are systematically underestimated.

The parts that the different types of valence electrons play in determining the magnetic structures of the heavy rare earth elements are thus clear; the itinerant *spd* electrons, common to all the heavy rare earth elements, mediate the interaction between magnetic moments and it is the nesting of their Fermi surfaces which can lead to instabilities in the paramagnetic phase with respect to the formation of incommensurate spin density waves. The *f* electrons, on the other hand, are responsible for setting up the magnetic moments and, as their number increases across the heavy rare earth series, they play an indirect role in promoting incommensurate order by shrinking the unit cell volume through the lanthanide contraction.

Received 31 July 2006; accepted 2 February 2007.

1. Gschneidner, K. A. & Eyring, L. (eds) *Handbook on the Physics and Chemistry of Rare Earths* Vol. 1 (North Holland, Amsterdam, 1978).
2. Jensen, J. & Mackintosh, A. K. *Rare Earth Magnetism* 286–304 (Clarendon, Oxford, 1991).
3. Keeton, S. C. & Loucks, T. L. Electronic structure of rare-earth metals. I. Relativistic augmented-plane-wave calculations. *Phys. Rev.* **168**, 672–678 (1968).
4. Fretwell, H. M. *et al.* Fermi surface as the driving mechanism for helical antiferromagnetic ordering in Gd-Y alloys. *Phys. Rev. Lett.* **82**, 3867–3870 (1999).
5. Cracknell, A. P. & Wong, K. C. *The Fermi Surface* (Clarendon, Oxford, 1973).
6. Taylor, K. N. R. & Darby, M. I. *Physics of Rare Earth Solids* 60–62 (Chapman and Hall, London, 1972).
7. Hubbard, J. Magnetism of iron. II. *Phys. Rev. B* **20**, 4584–4595 (1979).
8. Gyorffy, B. L., Pindor, A. J., Staunton, J., Stocks, G. M. & Winter, H. A first-principles theory of ferromagnetic phase transitions in metals. *J. Phys. F* **15**, 1337–1386 (1985).
9. Heinemann, M. & Temmerman, W. M. Magnetic structures of hcp bulk gadolinium. *Phys. Rev. B* **49**, 4348–4351 (1994).
10. Kurz, P., Bihlmayer, G. & Blügel, S. Magnetism and electronic structure of hcp Gd and the Gd(0001) surface. *J. Phys. Condens. Matter* **14**, 6353–6371 (2002).
11. Eriksson, O. *et al.* Bulk and surface magnetism and interplanar spacings in Gd from first-principles calculations. *Phys. Rev. B* **52**, 4420–4426 (1995).
12. Harmon, B. N., Antropov, V. P., Liechtenstein, A. I., Solovyev, I. V. & Anisimov, V. I. Calculation of magneto-optical properties for 4f systems: LSDA + Hubbard U results. *J. Phys. Chem. Solids* **56**, 1521–1524 (1994).
13. Perdew, J. P. & Zunger, A. Self-interaction correction to density-functional approximations for many-electron systems. *Phys. Rev. B* **23**, 5048–5079 (1981).
14. Strange, P., Svane, A., Temmerman, W. M., Szotek, Z. & Winter, H. Understanding the valency of rare earths from first-principles theory. *Nature* **399**, 756–758 (1999).
15. Lüders, M. *et al.* Self-interaction correction in multiple scattering theory. *Phys. Rev. B* **71**, 205109 (2005).
16. Rado, G. T. & Suhl, H. (eds) *Magnetism* Vol. IIB 337–377 (Academic Press, New York, 1966).
17. Staunton, J. B. & Gyorffy, B. L. Onsager cavity fields in itinerant-electron paramagnets. *Phys. Rev. Lett.* **69**, 371–374 (1992).
18. Turek, I., Kudrnovský, J., Bihlmayer, G. & Blügel, S. *Ab initio* theory of exchange interactions and the Curie temperature of bulk Gd. *J. Phys. Condens. Matter* **15**, 2771–2782 (2003).
19. Andrianov, A. V. Helical magnetic structures in heavy rare-earth metals as a probable manifestation of the electronic topological transition. *J. Magn. Magn. Mater.* **140–144**, 749–750 (1995).
20. Lifshitz, I. M. Anomalies of electron characteristics of a metal in the high pressure region. *Sov. Phys. JETP* **11**, 1130–1135 (1960).
21. Blanter, Y. M., Kaganov, M. I., Pantsulaya, A. V. & Varlamov, A. A. The theory of electronic topological transitions. *Phys. Rep.* **245**, 159–257 (1994).
22. Evenson, W. E. & Liu, S. H. Generalized susceptibilities and magnetic ordering of heavy rare earths. *Phys. Rev. Lett.* **21**, 432–434 (1968).
23. Dugdale, S. B. *et al.* Direct observation and calipering of the "webbing" Fermi surface of yttrium. *Phys. Rev. Lett.* **79**, 941–944 (1997).
24. Nordström, L. & Mavromaras, A. Magnetic ordering of the heavy rare earths. *Europhys. Lett.* **49**, 775–781 (2000).
25. Andrianov, A. V., Kosarev, D. I. & Beskrovnyi, A. I. Helical magnetic ordering in Tb completely suppressed by uniaxial tension: Evidence of electronic topological transition and support for the nesting hypothesis. *Phys. Rev. B* **62**, 13844–13847 (2000).
26. Milstein, F. & Robinson, L. B. Magnetic transitions in alloys of gadolinium and dysprosium. *Phys. Rev.* **159**, 466–472 (1967).
27. Andrianov, A. V. & Chistiakov, O. D. Evidence of pressure-induced antiferromagnetism in ferromagnetic $Ho_{0.4}Gd_{0.6}$. *Phys. Rev. B* **55**, 14107–14108 (1997).
28. Blundell, S. *Magnetism in Condensed Matter* 91–92 (Oxford Univ. Press, Oxford, 2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by the EPSRC (UK) and the CCLRC's Centre for Materials Physics and Chemistry. Computing resources were provided by the CSC at the University of Warwick, as well as the CCLRC's e-Science facility and the John von Neumann Institute for Computing in Jülich.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to I.D.H. (i.d.hughes@warwick.ac.uk) or J.B.S. (j.b.staunton@warwick.ac.uk).

LETTERS

Effect of evaporite deposition on Early Cretaceous carbon and sulphur cycling

Ulrich G. Wortmann[1] & Boris M. Chernyavsky[1]

The global carbon and sulphur cycles are central to our understanding of the Earth's history, because changes in the partitioning between the reduced and oxidized reservoirs of these elements are the primary control on atmospheric oxygen concentrations. In modern marine sediments, the burial rates of reduced carbon and sulphur are positively coupled, but high-resolution isotope records indicate that these rates were inversely related during the Early Cretaceous period[1]. This inverse relationship is difficult to reconcile with our understanding of the processes that control organic matter remineralization and pyrite burial. Here we show that the inverse correlation can be explained by the deposition of evaporites during the opening of the South Atlantic Ocean basin. Evaporite deposition can alter the chemical composition of sea water[2,3], which can in turn affect the ability of sulphate-reducing bacteria to remineralize organic matter and mediate pyrite burial. We use a reaction–transport model to quantify these effects, and the resulting changes in the burial rates of carbon and sulphur, during the Early Cretaceous period. Our results indicate that deposition of the South Atlantic evaporites removed enough sulphate from the ocean temporarily to reduce biologically mediated pyrite burial and organic matter remineralization by up to fifty per cent, thus explaining the inverse relationship between the burial rates of reduced carbon and sulphur during this interval. Furthermore, our findings suggest that the effect of changing seawater sulphate concentrations on the marine subsurface biosphere may be the key to understanding other large-scale perturbations of the global carbon and sulphur cycles.

The global cycling of sulphur (S) is controlled by the balance of S entering the ocean as a result of weathering, volcanic activity and hydrothermal fluxes, and the removal of S from the ocean in the form of S-bearing sediments. Oxidation of reduced S species consumes oxygen and forms dissolved sulphate, which is subsequently returned to the sedimentary reservoir in its oxidized form as $CaSO_4$, or it is reduced by bacterial activity and exported as pyrite. An increase of the pyrite oxidation rate relative to the pyrite burial rate therefore results in the net loss of atmospheric oxygen, and vice versa. A similar reasoning applies to carbon (C), which enters the sedimentary reservoir either as oxidized C (carbonate) or as reduced C (organic matter). However, sulphate reduction (pyrite production) is achieved by a microbially mediated redox reaction requiring the presence of organic matter as an electron donor. Because the capacity of most marine sediments to form pyrite is limited by organic-matter availability and not sulphate supply[4], perturbations to either the reduced C or reduced S flux are positively coupled to the other. As the microbially mediated reduction of C and S involves a considerable negative isotope effect (up to −70‰ for S, and about −28‰ for C; refs 5–8), an increase in the burial rate of the reduced species relative to the oxidized species results in a positive shift of the isotopic ratio, and vice versa. This allows us to trace the behaviour of the C and S cycles through Earth's history.

High-resolution $\delta^{34}S$ and $\delta^{13}C$ records[1] show that during the Early Cretaceous a large negative $\delta^{34}S$ excursion is accompanied by a large positive $\delta^{13}C$ excursion[9,10]. This implies that decreased pyrite burial rates occur together with increased organic-matter burial rates (we note that it is irrelevant whether this was caused by increased organic-matter production or increased organic-matter preservation). Previous studies explained this paradox by suggesting a global iron deficiency[11], or a global shift of organic-matter burial from the ocean to sulphate-limited continental settings[4]. Both scenarios seem unlikely during the Early Cretaceous, because the existing marine data clearly show increased organic-matter content or even black shale deposition[12,13], and there is little evidence to support reduced iron availability, or a shift of organic-matter deposition to terrestrial environments. A recent study suggested changes in volcanic and hydrothermal activity, together with increased weathering and reduced pyrite burial fluxes[1], to increase the flux of isotopically light S. However, this would also increase the flux of isotopically light C, and thus result in a positive coupling of the $\delta^{34}S$ and $\delta^{13}C$ signals.

Although it is difficult to explain the Early Cretaceous C–S paradox while assuming a modern ocean chemistry, we note that the Early Cretaceous perturbations of the C and S cycles are contemporaneous with the early stages of the opening of the South Atlantic. The break-up between South Africa and South America resulted in the creation of a silled 10^6 km^2 below-sealevel basin, which existed between 126 and 116 Myr ago[14]. During this time interval, evaporites totalling a volume of $1–4 \times 10^{15}$ m^3 were deposited in this basin[14,15]. Assuming that 20% of these evaporites consist of $CaSO_4$ (refs 16, 17), these deposits contain between 4.4×10^{18} to 17.6×10^{18} mol $CaSO_4$, equivalent to 26% to 100% of the current seawater sulphate reservoir. Here, we will therefore explore the impact of the South Atlantic evaporite sequences on the global cycling of S and C. We do this by coupling a sediment diagenetical model that describes organic-matter remineralization and pyrite burial as a function of seawater sulphate concentrations to a box model describing the global C and S cycles (see Methods and the Supplementary Information).

Our model captures the magnitude and timing of the marine $\delta^{13}C$ and $\delta^{34}S$ signals quite well (Figs 1 and 2). However, the magnitude of the $\delta^{13}C$ signal depends on the initial marine sulphate concentration: we obtain a positive shift of 3.5‰ assuming a 5 mM ocean, and +2.5‰ shift for 8 mM or 12 mM sulphate concentrations in the ocean (Figs 1 and 2). Depending on the input parameters chosen (an evaporite volume of 4×10^{15} m^3 and a 5 mM sulphate concentration, versus an evaporite volume of 1×10^{15} m^3 and 12 mM sulphate concentration), we obtain minimum or maximum estimates suggesting that the South Atlantic evaporites contain between 8% to 63% $CaSO_4$. These numbers are not too different from the long-term average gypsum content of evaporitic sequences (20%)[16,17], and may be affected by the high-calcium and low-sulphate concentrations of

Geobiology Isotope Laboratory, Department of Geology, University of Toronto, 22 Russell Street, Toronto, Ontario, M5S 3B1, Canada.

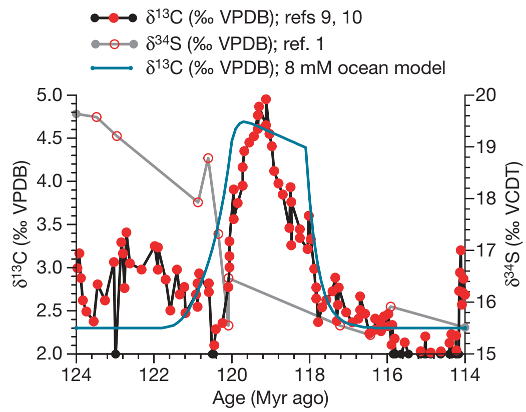


Figure 1 | A comparison between the observed[9,10] δ^{13}C and the δ^{13}C predicted by our C and S model and the δ^{34}S data[1]. Both magnitude and timing of the Aptian δ^{13}C excursion are well captured by our model. VCDT, Vienna Canyon Diablo Troilite. The numerical ages of the δ^{13}C data are based on ref. 27.

Early Cretaceous ocean water, which favour the precipitation of gypsum over other evaporitic phases[18].

The post-event shape of the modelled S-isotope data shows a first-order match with the published data (Fig. 2). However, the post-event isotope data depends on the choice of unconstrained post-event evaporite burial rates. It is therefore not possible to decide whether the second-order features visible in the published data are positive or negative excursions from the first-order signal. However, we tentatively relate the first and smallest excursion (112 Myr ago) to the onset of pyrite burial after marine sulphate levels increased sufficiently to sustain significant rates of organic-matter remineralization. The second and largest S-isotope excursion (95 Myr ago) is possibly related to increased pyrite burial rates during the Cenomanian black shale interval (OAE 2)[12,13,19]. However, these second-order features may also be explained by changes in weathering, volcanic and hydrothermal fluxes or changes in sea level[20–22].

The proposed dramatic changes in the marine sulphate budget must have left their mark on Early Cretaceous sedimentary sequences. We therefore investigated a chemostratigraphically well-dated shale section from the northern Tethyan margin in southern Germany[20,23]. These shales are characterized by a striking black/green cyclicity that records changes in organic-matter content (up to 5%), possibly related to astronomical forcing[23]. The short-term variability in organic-matter content is complemented by a remarkable long-term stability of the overall composition from Early Aptian to Late Albian times. That is, Fe and organic-matter contents do change from a black to a green layer, but if we compare layers of similar colour, major-element and trace-element ratios remain stable. The major exception to this pattern is the S content of the shales, which drops during the Early Aptian within a few million years by an order of magnitude from 0.5 to 0.05 wt%, and remains low for approximately three million years (Fig. 3). This drop is remarkable because it is unrelated to the organic-matter and Fe contents of the sediment. However, the timing of this dramatic reduction of sedimentary S content is coeval with the deposition of evaporites in the South Atlantic.

From these results and the range of published data on the South Atlantic evaporite sequences and Early Cretaceous sulphate concentrations, we conclude that large-scale evaporite depositional events can greatly alter the sulphate concentration of ocean water. These changes can affect the ability of the marine subsurface biosphere to remineralize organic matter and to mediate pyrite formation. This has several important consequences: (1) increased organic-matter burial rates may result in increased CO_2 drawdown rates; (2) decreased organic-matter remineralization rates must affect the global availability of phosphorus; (3) the drastically reduced sulphate concentrations must affect the ability of sulphate-reducing bacteria to oxidize methane, and thus modulate the global methane flux.

Our findings demonstrate that despite the different residence times of C and S in the ocean (10^3 versus 10^7 years), transient perturbations in either cycle must be either positively or negatively coupled to perturbations in the other, the respective magnitudes being a function of seawater sulphate concentrations. Low-sulphate concentrations dominated over most of the last 550 Myr (refs 18, 24,

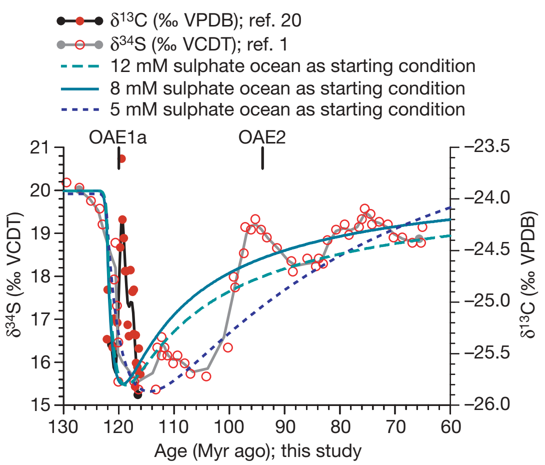


Figure 2 | Comparison of the published δ^{34}S data[1] and the δ^{13}C data[20] with three different model runs. The dotted line shows the results for a model assuming an initial marine sulphate concentration of 5 mM and a post-event decrease in sulphate burial rate. The solid line shows the result for an ocean with an initial sulphate concentration of 8 mM, in which the post-event sulphate burial rate is similar to the pre-event sulphate burial rate. The dashed line is similar to the 8 mM model, but assumes an ocean with an initial sulphate concentration of 12 mM. Circles represent data points, and the grey line represents a three-point moving average. OAE1a and OAE2 are major oceanic anoxic events.

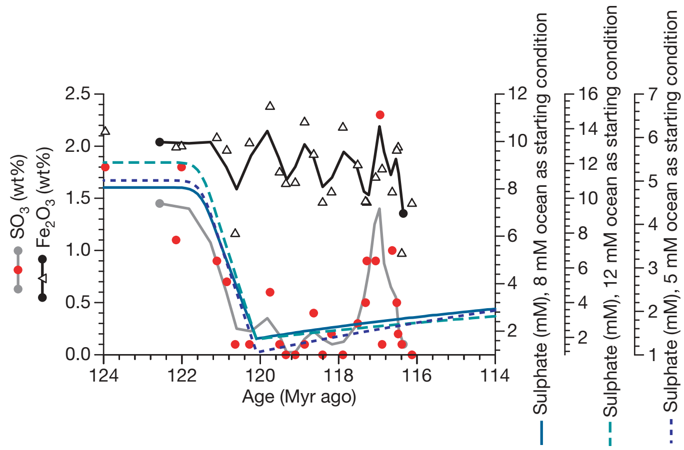


Figure 3 | Comparison between sedimentary S and Fe contents in an Aptian deep-sea section with model predicted seawater concentrations. In all model runs, the measured S content of the sediments mirrors the sharp decrease of the seawater concentrations. We note that the decreasing S values are not related to a decrease in Fe content. The decoupling between Fe and S is an artefact of the deep marine nature of the sampling site. organic matter in deep marine settings is highly refractory and reduction (and thus FeS_2 formation) is slow, continuing for millions of years after initial deposition (see ref. 28). Reduced diffusive fluxes therefore greatly alter the location where reduction takes place, shifting the location of FeS_2 formation. The fluctuations seen in the Fe data are caused by changes in bottom-water oxygenation[23]. The thick solid lines are obtained by processing the original data (circles and triangles) with a three-point moving-average window.

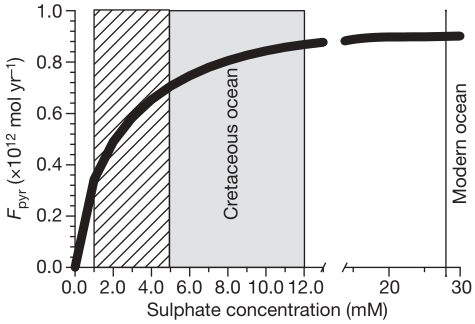


Figure 4 | Modelled response of the depth-integrated pyrite burial flux in an arbitrary sediment to changing sulphate concentrations at the sediment–water interface. The grey area shows the range of the published concentration values for Cretaceous ocean water. The hatched area shows the modelled seawater concentration range after the deposition of the South Atlantic evaporite sequences. Note the interrupted sulphate concentration scale.

25), so this interpretation may well be the key to understanding what has been termed the central dilemma of C–S geochemistry[3].

METHODS

Modelled seawater sulphate concentrations. To characterize the response of the marine subsurface biosphere to changing seawater sulphate concentrations, we use REMAP[26] to solve a reaction transport model describing the sulphate-concentration-dependent bacterial activity using the seawater sulphate concentration as a Dirichlet boundary condition at the upper boundary, and a zero-gradient Neumann condition at the lower boundary. Using this model, we calculate pyrite burial rates as a function of seawater sulphate concentrations (see Fig. 4). The resulting pyrite burial fluxes are scaled to equal the global flux at the given starting conditions of the model. The sulphate-concentration-dependent burial fluxes are now used as input parameters in a box model describing the evolution of the marine δ^{34}S and δ^{12}C values. For the C-cycle model, we assume as a first-order approximation that the remineralization rate is proportional to the pyrite burial rate, which depends on sulphate. We can thus formulate a time-dependent box model in which we relate organic-matter burial to sulphate availability.

The box model. For the box model, we assume an ocean volume of 1.38×10^{18} m^3 (see ref. 14) for all model runs, and present our results for an ocean with initial (that is, pre-event) sulphate concentrations of 5 mM, 8 mM and 12 mM covering the range of the published data[18]. Changes in the input fluxes of the S cycle have already been modelled[1], so we consider the hydrothermal, volcanic and weathering fluxes as a single and constant source, using a net input flux of 2×10^{12} mol yr^{-1} (ref. 1). Using these parameters, the model achieves steady-state conditions for an evaporitic output flux of 1.1×10^{12} mol yr^{-1} and a pyritic output flux of 0.90×10^{12} mol yr^{-1}. For the isotopic calculation, we use a seawater δ^{34}S of 20‰ (ref. 1), and a δ^{34}S of 5.76‰ for the combined hydrothermal, erosional and volcanic fluxes. The model achieves isotopic equilibrium for a seawater-pyrite difference of −31.55‰.

The C-cycle model. The C-cycle model assumes that the mass of dissolved C in the ocean is 3.2×10^{18} mol, and a net input flux of 25.43×10^{12} mol C yr^{-1} with a δ^{13}C value of −5.98‰ VPDB (the Vienna Pee-Dee Belemnite standard). The output flux into the oxidized reservoir is kept constant at 20×10^{12} mol yr^{-1} and is assumed to happen in isotopic equilibrium. The initial burial flux of organic matter is set at 5.43×10^{12} mol yr^{-1} with an isotopic offset of −28‰ relative to the contemporaneous sea water. The combined C- and S-cycle model is then run by perturbing the steady state with a pulsed evaporitic depositional event, where the size of this event is modified until a good fit between predicted and measured S data is achieved. A detailed discussion of the above models and their parametrization is given in the Supplementary Information.

Received 7 April 2006; accepted 21 February 2007.

1. Paytan, A., Kastner, M., Campbell, D. & Thiemens, M. H. Seawater sulfur isotope fluctuations in the Cretaceous. *Science* **304**, 1663–1665 (2004).
2. Arthur, M. A. in *Climate in Earth History* 55–67 (National Academy Press, Washington DC, 1982).
3. Holser, W. T., Schidlowski, M., Mackenzie, F. T. & Maynard, J. B. in *Chemical Cycles in the Evolution of the Earth* 105–174 (John Wiley & Sons, New York, 1988).
4. Berner, R. A. & Raiswell, R. Burial of organic carbon and pyrite sulfur in sediments over phanerozoic time: a new theory. *Geochim. Cosmochim. Acta* **47**, 855–862 (1983).
5. Wortmann, U. G., Bernasconi, S. M. & Böttcher, M. E. Hypersulfidic deep biosphere indicates extreme sulfur isotope fractionation during single step microbial sulfate reduction. *Geology* **29**, 647–650 (2001).
6. Brunner, B. & Bernasconi, S. M. A revised isotope fractionation model for dissimilatory sulfate reduction in sulfate reducing bacteria. *Geochim. Cosmochim. Acta* **69**, 4759–4771 (2005).
7. Deines, P. in *Handbook of Environmental Geochemistry* (eds Fritz, P. & Fontes, J. C.) Vol. 1, 239–406 (Elsevier, New York, 1980).
8. O'Leary, M. H. Carbon isotope fractionation in plants. *Phytochemistry* **20**, 553–567 (1981).
9. Herrle, J. O., Kößler, P., Friedrich, O., Erlenkeuser, H. & Hemleben, C. High-resolution carbon isotope records of the Aptian to Lower Albian from SE France and the Mazagan Plateau (DSDP Site 545): a stratigraphic tool for paleoceanographic and paleobiologic reconstruction. *Earth Planet. Sci. Lett.* **218**, 149–161 (2004).
10. Erba, E. *et al.* Integrated stratigraphy of the Cismon Apticore (Southern Alps, Italy): A 'reference section' for the Hautrivian-Aptian interval at low latitudes. *J. Foraminiferal Res.* **29**, 371–391 (1999).
11. Raiswell, R. & Canfield, D. E. Sources of iron for pyrite formation in marine sediments. *Am. J. Sci.* **298**, 219–245 (1988).
12. Ryan, W. B. F. & Cita, M. B. Ignorance concerning episodes of ocean-wide stagnation. *Mar. Geol.* **23**, 197–215 (1977).
13. Schlanger, S. O. & Jenkyns, H. C. Cretaceous oceanic anoxic events: causes and consequences. *Geol. Mijnbouw* **55**, 179–184 (1976).
14. Burke, K. & Şengör, C. Ten metre global sea-level change associated with South Atlantic Aptian salt deposition. *Mar. Geol.* **83**, 309–312 (1988).
15. Southam, J. R. & Hay, W. W. in *The Oceanic Lithosphere* (ed. Emiliani, E.) Vol. 7 of *The Sea* 1617–1684 (Wiley-Interscience, New York, 1981).
16. Zharkov, M. A. *History of Paleozoic Salt Accumulation* 1–308 (Springer, New York, 1981).
17. Rhonov, A. B. The earth's sedimentary shell (quantitative patterns of its structure, compositions, and evolution). *Int. Geol. Rev.* **24**, 1365–1388 (1982).
18. Lowenstein, T. K., Hardie, L. A., Timofeef, M. N. & Demicco, R. V. Secular variation in seawater chemistry and the origin of calcium chloride basinal brines. *Geology* **31**, 857–860 (2003).
19. Arthur, M. A. & Premoli Silva, I. in *Nature and Origin of Cretaceous Carbon-Rich Facies* (eds Schlanger, S. O. & Cita, M. B.) 7–54 (Academic, San Diego, California, 1982).
20. Wortmann, U. G., Herrle, J. O. & Weissert, H. Altered carbon cycling and coupled changes in Early Cretaceous weathering patterns: Evidence from integrated carbon isotope and sandstone records of the western Tethys. *Earth Planet. Sci. Lett.* **220**, 69–82 (2004).
21. Weissert, H. & Erba, E. Volcanism, CO_2 and paleoclimate: A Late Jurassic-Early Cretaceous carbon and oxygen isotope record. *J. Geol. Soc.* **161**, 695–702 (2004).
22. Bjerrum, C. J., Bendtsen, J. & Legarth, J. J. F. Modeling organic carbon burial during sea level rise with reference to the Cretaceous. *Geochem. Geophys. Geosyst.* **7**, Q05008, doi:10.1029/2005GC001032 (2006).
23. Wortmann, U. G., Zacher, W. & Hesse, R. Major-element data from cyclic black shales and their paleoceanographic implications for the early Cretaceous deep western Tethys. *Paleoceanography* **14**, 525–541 (1999).
24. Berner, R. A. A model for calcium, magnesium and sulfate in seawater over phanerozoic time. *Am. J. Sci.* **304**, 438–453 (2004).
25. Holland, H. D. Sea level, sediments and the composition of seawater. *Am. J. Sci.* **305**, 220–239 (2005).
26. Chernyavsky, B. & Wortmann, U. G. REMAP: A reaction transport model for isotope ratio calculations in porous media. *Geochem. Geophys. Geosyst.* **8**, Q02009, doi:10.1029/2006GC001442 (2006).
27. Leckie, R. M., Bralower, T. J. & Cashman, R. Oceanic anoxic events and plankton evolution: Biotic response to tectonic forcing during the mid-Cretaceous. *Paleoceanography* **17**, 1–29 doi:10.1029/2001PA000623 (2002).
28. Parkes, R. J. *et al.* Deep sub-seafloor prokaryotes stimulated at interfaces over geological time. *Nature* **436**, 390–394 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. Kastner, B. Hay, H. Weissert, N. Andersen, A. J. M. Stams and B. Brunner for discussions. J. Bollmann, E.T.C. Spooner, J. Walker, L. Lee, J. Yap and C. Banks provided helpful comments on an earlier version of this manuscript. We gratefully acknowledge the comments of M. Arthur. M. Gorton, H. Li and C. Greyrobe provided assistance in the laboratory. C. Marra participated in this study as part of a University of Toronto mentorship program for gifted high school students. This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC).

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to U.G.W. (uli.wortmann@utoronto.ca).

Vol 446|5 April 2007|doi:10.1038/nature05667 nature

LETTERS

Geomagnetic field strength 3.2 billion years ago recorded by single silicate crystals

John A. Tarduno[1,2], Rory D. Cottrell[1], Michael K. Watkeys[3] & Dorothy Bauch[1]

The strength of the Earth's early geomagnetic field is of importance for understanding the evolution of the Earth's deep interior, surface environment and atmosphere. Palaeomagnetic and palaeointensity data from rocks formed near the boundary of the Proterozoic and Archaean eons, some 2.5 Gyr ago, show many hallmarks of the more recent geomagnetic field. Reversals are recorded[1], palaeosecular variation data[2] indicate a dipole-dominated morphology and available palaeointensity values are similar to those from younger rocks[1–3]. The picture before 2.8 Gyr ago is much less clear. Rocks of the Archaean Kaapvaal craton (South Africa) are among the best-preserved, but even they have experienced low-grade metamorphism[4]. The variable acquisition of later magnetizations by these rocks is therefore expected, precluding use of conventional palaeointensity methods. Silicate crystals from igneous rocks, however, can contain minute magnetic inclusions capable of preserving Archaean-age magnetizations. Here we use a CO_2 laser heating approach and direct-current SQUID magnetometer measurements to obtain palaeodirections and intensities from single silicate crystals that host magnetite inclusions. We find 3.2-Gyr-old field strengths that are within 50 per cent of the present-day value, indicating that a viable magnetosphere sheltered the early Earth's atmosphere from solar wind erosion.

Néel's[5] theory can be used to derive time–temperature relationships that are useful for predicting the acquisition of thermoviscous magnetization. According to this theory, the thermal relaxation time τ for single-domain magnetic grains can be expressed as[6]:

$$\frac{1}{\tau} = \frac{1}{\tau_0} \exp\left[-\frac{\mu_0 V M_s H_K}{2kT}\left(1 - \frac{|H_0|}{H_K}\right)^2\right] \qquad (1)$$

where τ_0 (10^{-9} s) is the interval between thermal excitations, μ_0 is the permeability of free space, V is grain volume, M_s is spontaneous magnetization, k is Boltzmann's constant, T is temperature and H_0 is the applied field. The microcoercive force H_K measures the field needed to rotate the magnetization without thermal excitation[6]. Given $H_K \gg H_0$, and relaxation times τ_A and τ_B representing temperatures T_A and T_B, respectively, we can write[7]:

$$\frac{T_A \ln(\tau_A/\tau_0)}{M_s(T_A) H_K(T_A)} = \frac{T_B \ln(\tau_B/\tau_0)}{M_s(T_B) H_K(T_B)} \qquad (2)$$

This relationship suggests that, for low-grade metamorphism occurring over 1 Myr, single-domain magnetic grains with blocking temperatures less than 400 °C could be contaminated by overprints (see Supplementary Information). For multidomain magnetic grains, we can expect to measure partial thermoviscous magnetizations at unblocking temperatures up to the Curie point of magnetite[6].

With increased temperatures, there is also potential for the growth of new magnetic minerals. For example, palaeointensities 4 to 10 times lower than present-day values have been reported from 3.5-Gyr-old lavas (the Komati Formation) of the Barberton greenstone belt[8]. The magnetization was originally thought to be a thermoremanent magnetization related to early greenschist metamorphism, but a subsequent study concluded it to be a chemical remanent magnetization related to magnetite grain growth[9]. Chemical remanent magnetizations provide only minimum bounds on field strength. Moreover, the Barberton area has also seen low-grade metamorphism in late Archaean[4], and possibly Proterozoic, times. The path to magnetic grain growth in iron-rich komatiites during metamorphism is direct, and the chemical remanent magnetization could be much younger than the rock age. Directions from whole-rock samples of 3.45-Gyr-old lavas of the Pilbara craton (Australia) have also been interpreted as a sign of an early dynamo[10]. However, these data lack field strength estimates and may also be contaminated by secondary magnetizations related to later metamorphism (see Supplementary Information).

The challenge was to find an alternative magnetic recorder that can see through the complex set of overprinted secondary magnetizations that should affect Archaean rocks through thermochemical processes. The acquisition of secondary magnetizations is related to magnetic domain state, so single-domain or pseudo-single-domain carriers needed to be isolated. In addition, we looked for samples with a minimum of iron that might otherwise migrate to form new magnetic minerals during metamorphism.

Magnetic properties of individual rock-forming silicate grains have been shown to be a valuable means of assessing palaeointensity and directions[11–13] because these crystals can contain single-domain and near-single-domain (pseudo-single-domain) magnetic inclusions protected by the silicate host. We applied this approach to the study of the Dalmein and Kaap Valley Plutons that intrude the Barberton greenstone belt. These are granodioritic and tonalitic, respectively, and both have been dated to 3.2 Gyr ago by U–Pb geochronology[14]. We focused on feldspar (microcline), quartz and hornblende. To investigate the promise of these mineral carriers, we first evaluated their rock magnetic properties. Isothermal remanent acquisition curves do not indicate the presence of high-coercivity magnetic inclusions such as haematite. Low-temperature data collected using a Magnetic Properties Measurement System show the variable presence of the Verwey transition[6]—the cubic to monoclinic phase transition in magnetite—on cooling through 120 K (Supplementary Fig. 1). These data indicate that all three crystal types can have magnetite inclusions. However, magnetic hysteresis properties measured using a Princeton Measurements Corporation alternating gradient force magnetometer define different mean magnetic domain states of the silicate-hosted inclusions. Hornblende has inclusions with near multidomain behaviour, whereas inclusions in quartz and microcline have pseudo-single-domain to single-domain characteristics (Supplementary Fig. 2). Because of its high overprint potential, we excluded hornblende from further consideration. Quartz and microcline did not display

[1]Department of Earth and Environmental Sciences, [2]Department of Physics and Astronomy, University of Rochester, Rochester, New York 14627, USA. [3]School of Geological Sciences, University of KwaZulu-Natal, Durban 4041, South Africa.

significant magnetic anisotropies (Supplementary Fig. 3), and are also attractive because they contain minimal amounts of reactive iron.

To test whether the isolated remanence was far removed from later field directions, we developed a new CO_2 laser heating approach to link palaeointensity data from single silicate crystals to directional components (see Supplementary Information). Our procedure begins with a variation of micro-sampling[15] that uses oriented thin sections cut from palaeomagnetic cores. The sections are much thicker than those used for standard petrographic studies (>1 mm). After undesirable minerals (for example, hornblende) are mechanically etched away, the sections are cut into smaller subsections, each having a single mineralogy. Each subsection holds a single grain with a length of ~2–3 mm, which is used for directional and Thellier palaeointensity analyses.

Subsections are demagnetized using a 20 W Synrad CO_2 laser, which yields a 10.6 μm infrared beam that couples well with silicates. Renne and Onstott[16] previously used a ruby laser to selectively demagnetize grains by ablation. Our approach differs in that the CO_2 laser is used to heat the crystals gently. Heating is carried out on timescales of a few minutes, minimizing total heating duration and potential thermally induced alteration while facilitating the measurement of oriented samples.

Testing and calibration were performed using infrared cameras and silicate crystals separated from younger rocks for which directional data were available (Supplementary Fig. 4). All remanence measurements were made using a 2 G direct-current superconducting quantum interference device (SQUID) magnetometer with a high-resolution coil configuration.

Quartz and microcline single crystals have stable remanence components at high unblocking temperatures after magnetizations at lower unblocking temperatures are removed (Fig. 1). This pattern is consistent with the predicted acquisition of overprints by magnetic grains with low unblocking temperatures (equations (1) and (2)) and the preservation of primary signals by grains with higher unblocking temperatures.

Undeformed mafic dykes ranging from Archaean to Mesozoic in age in the Dalmein and Kaap Valley Plutons are vertical, suggesting that the plutons have not undergone significant tilting since emplacement. We can use magnetic measurements from these dykes to test whether the silicate crystals preserve Archaean directions. A dyke from the Kaap Valley Pluton yields two components of magnetization (Supplementary Fig. 5): a steep component at intermediate unblocking temperatures (~200 °C to 450 °C); and a shallow high-unblocking-temperature component (500 °C to 580 °C). Thermomagnetic experiments (Supplementary Fig. 6) and scanning electron microscopy suggest that the shallow component is carried by magnetite formed during cooling (that is, high-temperature oxidation). It is similar to directions reported from dykes of the 1.1-Gyr-old Umkondo large igneous province[17]. However, the steep direction is similar to ~2-Gyr-old overprints previously reported from the Kaapvaal craton[18]. Given the rock magnetic characteristics of the dyke, and its geologic setting, it is unlikely that the intermediate-unblocking-component of magnetization is older than the component isolated at higher unblocking temperatures. Therefore, the shallow component is either older than 2.0 Gyr, or the steep direction is a combination of several later magnetizations (and thus not a true reflection of the magnetic field 2.0 Gyr ago) (Fig. 2). A dyke from the Dalmein Pluton also carries a high-unblocking-temperature component, but its direction suggests the dyke is related to Karoo magmatism[19] 180 Myr ago.

The derived palaeomagnetic directions from the silicate minerals (Fig. 2): (1) agree between plutons; (2) differ from the directions of younger intruding dykes (Supplementary Tables 1 and 2); and (3) record dual polarity (Kaap Valley Pluton). The data from the Kaap Valley Pluton hint at some polarity asymmetry, but we cannot determine from the available data whether this reflects incomplete sampling of a time-varying field, or a longer-term non-dipole morphology. However, the directions preserved by the Kaap Valley Pluton quartz and feldspar clearly differ from those of a prior study of whole rocks[20], which were thought to record a pulse of rapid (>16 mm yr^{-1}) Archaean plate motion[21]. We feel that directions from the hornblende-rich whole rocks of the Kaap Valley Pluton instead reflect various extents of overprinting (equations (1) and (2)).

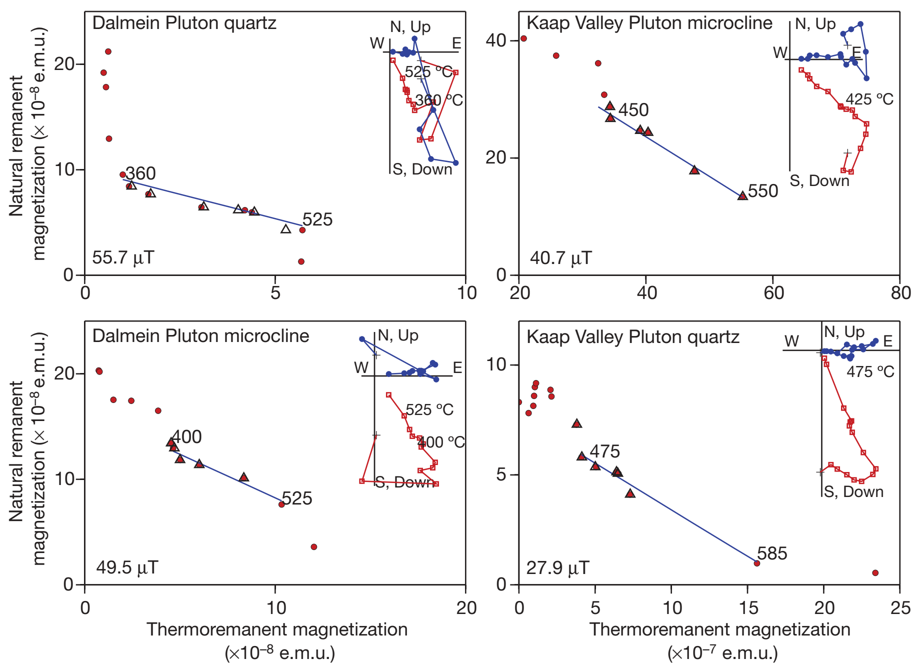


Figure 1 | Examples of natural remanent magnetization versus thermoremanent magnetization data from Thellier experiments on Archaean oriented single silicate crystals using a CO_2 laser/SQUID system. Triangles are partial-thermoremanent magnetization experimental checks (see ref. 12). The line is the least-squared fit to natural remanent magnetization versus thermoremanent magnetization data (circles); the resulting field estimate in μT is shown. Inset of each plot shows orthogonal vector plot of field-off steps used to constrain palaeomagnetic directions. Red squares, inclination; blue circles, declination.

Palaeointensity data that meet acceptance criteria[12] yield 21 raw field values between approximately 20 μT and 60 μT (Supplementary Tables 1 and 3). Because previous palaeointensity results from the Barberton komatiite lavas record a chemical remanent magnetization[9] of uncertain age (but potentially as young as 2.8 Gyr[22], Fig. 2), the new data represent the earliest direct measurement of Earth's magnetic field strength. There is agreement between values obtained from quartz and microcline, and between results obtained using laser and conventional heating (Supplementary Table 3). Together with the directional data, these values suggest virtual dipole moments of $7.5 \pm 1.2 \times 10^{22}$ A m^2 and $6.4 \pm 2.1 \times 10^{22}$ A m^2 for the Dalmein and Kaap Valley Plutons, respectively. Both estimates are similar to the present field strength.

Cooling rates in nature vary greatly from those of our experiments, so we examine cooling rate corrections[23] using geochronological and geological constraints (see Supplementary Information). These corrections suggest that the raw values could overestimate the strength by 47% to 57%. A limitation of the corrections is their reliance on theoretical considerations of elongate single-domain grains. The effect on pseudo-single-domain grains, for example, is likely to be much less[24]. Nevertheless, applying the corrections can yield useful lower bounds on field strength. The corrected virtual dipole moment from the Dalmein Pluton is $4 \pm 1 \times 10^{22}$ A m^2, whereas the corrected value from the Kaap Valley Pluton is $3 \pm 1 \times 10^{22}$ A m^2. We note that the lower bound from the Kaap Valley Pluton is within 50% of the value[25] thought to characterize the Brunhes chron (0–0.78 Myr ago), whereas the lower bound for the Dalmein Pluton value is within 50% of the intensity of the present-day field. The present-day dipole field is rapidly losing strength[26]; if it falls to half its current value, non-dipole foci will emerge, allowing deeper penetration of highly energetic solar particles into the atmosphere[27]. However, the overall field will remain relatively strong, maintaining a significant magnetosphere. Therefore, on the basis of the lower bounds of field strength provided by our new data, we expect that a magnetosphere would have sheltered the early Archaean Earth from solar-wind-related atmospheric erosion.

The geodynamo could have commenced more than 1 Gyr earlier than the new records available from the Kaapvaal craton silicate minerals, with formation of the liquid iron core shortly after planetary accretion. However, this contrasts with a recent inference based on the abundance and isotopic ratio of nitrogen in lunar soils. These suggest a ~3.8–3.9-Gyr-old terrestrial source, eroded from Earth's atmosphere by the solar wind[28] in the absence of a strong magnetic field and magnetosphere. If the hypothesis for the origin of lunar nitrogen is correct, our direct palaeo-field measurements constrain the start of the dynamo to between 3.9 and 3.2 Gyr ago. This is a later start than on Mars, where magnetic anomalies from terrains older than ~4 Gyr reflect a past dynamo[29]. Earth's dynamo may have started because the onset of solid inner-core growth resulted in a change from stratified to convective core flow. Recent considerations suggest that an inner-core age this old is compatible with available compositional and thermal constraints[30]. The smaller size and faster cooling of Mars relative to Earth, together with differences in its accretionary history, may have led to either an earlier initiation of inner-core growth, or direct stimulation of liquid-core convection. Either mechanism could have driven an earlier, but ultimately unsustainable, Martian dynamo.

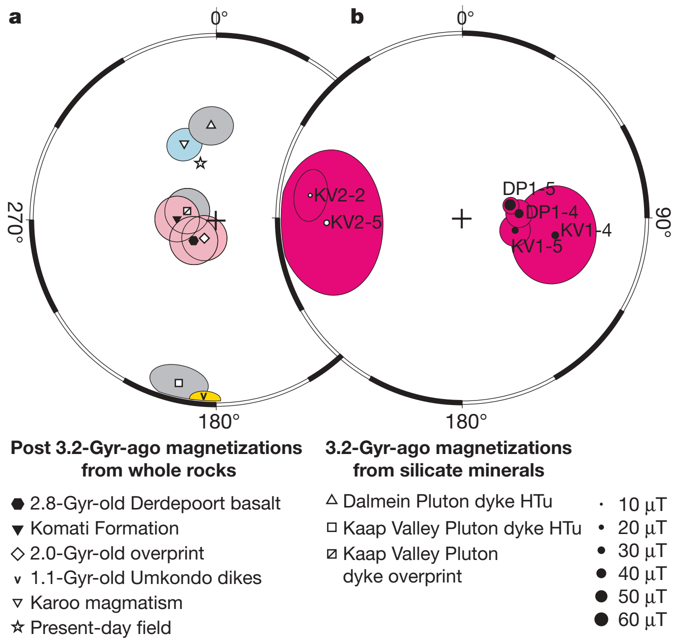


Figure 2 | Stereonets showing palaeomagnetic and palaeointensity results from the Kaapvaal craton, South Africa. Filled symbols, positive inclinations; open symbols, negative inclinations. **a**, Post 3.2-Gyr-ago directions from whole rocks with 95% confidence intervals (grey) with directions predicted for the Barberton area using Mesozoic–Archaean palaeomagnetic poles (colours are keyed to age: magenta, 3.2 Gyr; pink, Archaean–Proterozoic; yellow, Proterozoic; blue, Jurassic; grey, unknown age). HTu, high-unblocking-temperature component. The Kaap Valley Pluton dyke overprint is the component isolated at intermediate unblocking temperatures. Data sources for other directions: Derdepoort basalt, ref. 22; Komati Formation, ref. 9; the 2.0-Gyr-old overprint, ref. 18; Umkondo dykes, ref. 17; Karroo magmatism, ref. 19. **b**, Results from oriented quartz and microcline crystals. Symbol size represents intensity; 95% confidence regions around the mean direction are also shown (magenta). DP, Dalmein Pluton silicate crystal (quartz and feldspar) means; KV, Kaap Valley Pluton silicate crystal (quartz and feldspar) means. Labels refer to samples/sites (see Supplementary Information Tables).

Received 30 November 2006; accepted 7 February 2007.

1. Dunlop, D. J. & Yu, Y. in *Timescales of the Paleomagnetic Field* (eds Channell, J. E. T., Kent, D. V., Lowrie, W. & Meert, J. G) 85–100 (Geophys. Monograph Ser. 145, American Geophysical Union, Washington DC, 2004).
2. Smirnov, A. V. & Tarduno, J. A. Secular variation of the Late Archean–Early Proterozoic geodynamo. *Geophys. Res. Lett.* **31,** L16607 (2004).
3. Smirnov, A. V., Tarduno, J. A. & Pisakin, B. N. Paleointensity of the Early Geodynamo (2.45 Ga) as recorded in Karelia: a single crystal approach. *Geology* **31,** 415–418 (2003).
4. Tice, M. M., Bostick, B. C. & Lowe, D. R. Thermal history of the 3.5–3.2 Ga Onverwacht and Fig Tree Groups, Barberton greenstone belt, South Africa, inferred by Raman microspectroscopy of carbonaceous material. *Geology* **32,** 37–40 (2004).
5. Néel, L. Some theoretical aspects of rock magnetism. *Adv. Phys.* **4,** 191–243 (1955).
6. Dunlop, D. J. & Özdemir, Ö. *Rock Magnetism, Fundamentals and Frontiers* (Cambridge Univ. Press, Cambridge, UK, 1997).
7. Pullaiah, G., Irving, E., Buchan, K. L. & Dunlop, D. J. Magnetization changes caused by burial and uplift. *Earth Planet. Sci. Lett.* **28,** 133–143 (1975).
8. Hale, C. J. The intensity of the geomagnetic field at 3.5 Ga: Paleointensity results from the Komati Formation, Barberton Mountain Land, South Africa. *Earth Planet. Sci. Lett.* **86,** 354–364 (1987).
9. Yoshihara, A. & Hamano, Y. Paleomagnetic constraints on the Archean geomagnetic field intensity obtained from komatiites of the Barberton and Belingwe greenstone belts, South Africa and Zimbabwe. *Precambr. Res.* **131,** 111–142 (2004).
10. McElhinny, M. W. & Senanayake, W. E. Paleomagnetic evidence for the existence of the geomagnetic field at 3.5 Ga Ago. *J. Geophys. Res.* **85,** 3523–3528 (1980).
11. Cottrell, R. D. & Tarduno, J. A. Geomagnetic paleointensity derived from single plagioclase crystals. *Earth Planet. Sci. Lett.* **169,** 1–5 (1999).
12. Cottrell, R. D. & Tarduno, J. A. In search of high fidelity geomagnetic paleointensities: A comparison of single crystal and whole rock Thellier-Thellier analyses. *J. Geophys. Res.* **105,** 23,579–23,594 (2000).
13. Tarduno, J. A., Cottrell, R. D. & Smirnov, A. V. The paleomagnetism of single silicate crystals: Recording geomagnetic field strength during mixed polarity intervals, superchrons and inner core growth. *Rev. Geophys.* **44,** RG1002, doi:10.1029/2005RG000189 (2006).
14. Poujol, M., Robb, L. J., Anhaeusser, C. R. & Gericke, B. A review of the geochronological constraints on the evolution of the Kaapvaal Craton, South Africa. *Precambr. Res.* **127,** 181–213 (2003).
15. Geissman, J. W., Harlan, S. S. & Brearley, A. J. The physical isolation and identification of carriers of geologically stable remanent magnetization: Paleomagnetic and rock magnetic microanalysis and electron microscopy. *Geophys. Res. Lett.* **15,** 479–482 (1988).

16. Renne, P. R. & Onstott, T. C. Laser-selective demagnetization: A new technique in paleomagnetism and rock magnetism. *Science* 242, 1152–1155 (1988).
17. Gose, W. A., Hanson, R. E., Dalziel, I. W. D., Pancake, J. A. & Seidel, E. K. Paleomagnetism of the 1.1 Ga Umkondo large igneous province in southern Africa. *J. Geophys. Res.* 111, B09101 (2006).
18. Layer, P. W., Lopez-Martinez, M., Kroner, A., York, D. & McWilliams, M. Thermochronometry and palaeomagnetism of Archaean Nelshoogte Pluton, South Africa. *Geophys. J. Int.* 135, 129–145 (1998).
19. Hargraves, R. B., Rehacek, J. & Hooper, P. R. Palaeomagnetism of the Karoo igneous rocks in southern Africa. *S. Afr. J. Geol.* 100, 195–212 (1997).
20. Layer, P. W., Kroner, A. & McWilliams, M. An Archean geomagnetic reversal in the Kaap Valley pluton, South Africa. *Science* 273, 943–946 (1996).
21. Kroner, A. & Layer, P. W. Crust formation and plate motion in the early Archean. *Science* 256, 1405–1411 (1992).
22. Wingate, M. T. D. A palaeomagnetic test of the Kaapvaal-Pilbara (Vaalbara) connection at 2.78 Ga. *S. Afr. J. Geol.* 101, 257–274 (1998).
23. Halgedahl, S. L., Day, R. & Fuller, M. The effect of cooling rate on the intensity of weak-field TRM in single-domain magnetite. *J. Geophys. Res.* 85, 3690–3698 (1980).
24. McClelland Brown, E. Experiments on TRM intensity dependence on cooling rate. *Geophys. Res. Lett.* 11, 205–208 (1984).
25. Valet, J. P. Time variations in geomagnetic intensity. *Rev. Geophys.* 41, 1004 (2003).
26. Hulot, G., Eymin, C., Langlais, B., Mandea, M. & Olsen, N. Small-scale structure of the geodynamo inferred from Oersted and Magsat satellite data. *Nature* 416, 620–623 (2002).
27. Sinnhuber, M. *et al.* A model study of the impact of magnetic field structure on atmospheric composition during solar proton events. *Geophys. Res. Lett.* 30, 1818 (2003).
28. Ozima, M. *et al.* Terrestrial nitrogen and noble gases in lunar soils. *Nature* 436, 655–659 (2005).
29. Stevenson, D. J. Mars' core and magnetism. *Nature* 412, 214–219 (2001).
30. Gubbins, D., Alfe, D., Masters, G., Price, G. D. & Gillan, M. Gross thermodynamics of two-component core convection. *Geophys. J. Int.* 157, 1407–1414 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank P. Layer, A. Smirnov and P. Doubrovine for discussions. This research was supported by NSF.

Author Contributions J.A.T., M.K.W. and D.B. took part in field studies, and J.A.T., R.D.C. and D.B. in technique development, experiments and data analysis.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.A.T. (john@earth.rochester.edu).

Vol 446 | 5 April 2007 | doi:10.1038/nature05682

LETTERS

Doushantuo embryos preserved inside diapause egg cysts

Leiming Yin[1], Maoyan Zhu[1], Andrew H. Knoll[2], Xunlai Yuan[1], Junming Zhang[1] & Jie Hu[1]

Phosphatized microfossils in the Ediacaran (635–542 Myr ago) Doushantuo Formation, south China, have been interpreted as the embryos of early animals[1–4]. Despite experimental demonstration that embryos can be preserved[5], microstructural evidence that the Doushantuo remains are embryonic[6] and an unambiguous record of fossil embryos in Lower Cambrian rocks[7], questions about the phylogenetic relationships of these fossils remain. Most recently, some researchers have proposed[8] that Doushantuo microfossils may be giant sulphur-oxidizing bacteria comparable to extant *Thiomargarita* sp. Here we report new observations that provide a test of the bacterial hypothesis. The discovery of embryo-like Doushantuo fossils inside large, highly ornamented organic vesicles (acritarchs) indicates that these organisms were eukaryotic, and most probably early cleavage stage embryos preserved within diapause egg cysts. Large acanthomorphic microfossils of the type observed to contain fossil embryos first appear in rocks just above a 632.5 ± 0.5-Myr-old ash bed[9], suggesting that at least stem-group animals[6] inhabited shallow seas in the immediate aftermath of global Neoproterozoic glaciation.

Simple embryos and *Thiomargarita* sp. are both spheroidal, and both divide repeatedly without intervening growth to form multicellular structures with overall dimensions comparable to those of the single cells from which they derive. Although members of the genus *Thiomargarita* lack a number of features observed in Doushantuo fossils[10,11], it is later developmental stages that would provide definitive support for the embryo hypothesis; however, these have been recognized only tentatively[11]. There is another test, though. Bacteria may be surrounded by polysaccharide envelopes of simple morphology, but they do not form large, preservable cysts with complex ornamentation. In contrast, the life cycles of many eukaryotes, including some animals, include resting stages characterized by ornate and preservable walls[12]. Doushantuo beds containing fossils interpreted as embryos also preserve abundant and diverse acritarchs characterized by large size (100 to ~700 μm diameter) and conspicuous ornamentation that is distributed symmetrically over the vesicle surface (Fig. 1)[13,14]. Among the extant organisms, by far the closest morphological matches to these fossils are diapause egg cysts, or hulls, produced by a phylogenetically disparate array of aquatic invertebrates[15,16].

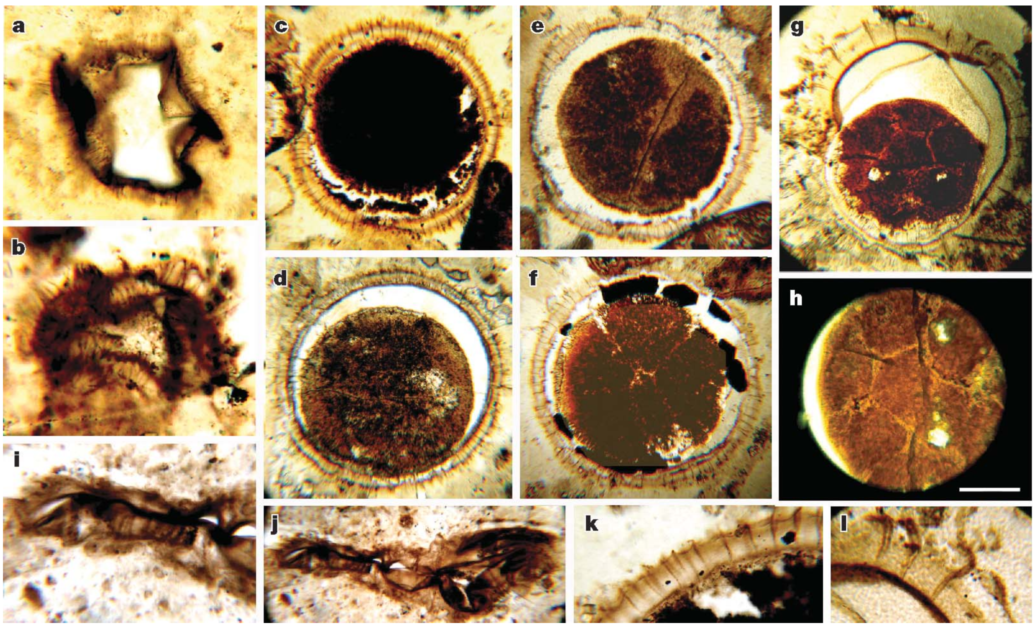


Figure 1 | Large acanthomorphic acritarchs, some containing preserved embryos, in the Ediacaran Doushantuo Formation, Xiaofenghe section, western Hubei Province, China. **a**, *Ericiasphaera magna*, sample number XFHB-35. **b**, *Appendisphaera tenuis*, sample number XFHB-33. **c**–**l**, *Tianzhushania spinosa*; **c**–**h**, **k**, **l** sample number XFHB-2; **i**, **j** sample number XFHB-29. Panels **c**–**h** illustrate one (**c**), two (**d**, **e**), eight (**f**) and 16 (**g**, with greater detail in **h**) cell stages of preserved embryos within encompassing structures interpreted as diapause egg cysts; **k**, **l**, show details of wall structure in the specimens illustrated in **c** and **g**. Panel **j**, magnified in **l**, shows a *Tianzhushania* fossil from ~ 632-Myr-old beds of the lower Doushantuo Formation. Scale bar (in h) represents 32 μm for **a**; 30 μm for **b**; 185 μm for **c**; 182 μm for **d**; 185 μm for **e**; 163 μm for **f**; 195 μm for **g**; 130 μm for **h**; 50 μm for **i**; 98 μm for **j**; 38 μm for **k**; and 98 μm for **l**.

[1]State Key Laboratory of Paleobiology and Stratigraphy, Nanjing Institute of Geology and Palaeontology, Chinese Academy of Sciences, Nanjing 210008, China. [2]Botanical Museum, Harvard University, Cambridge, Massachusetts 02138, USA.

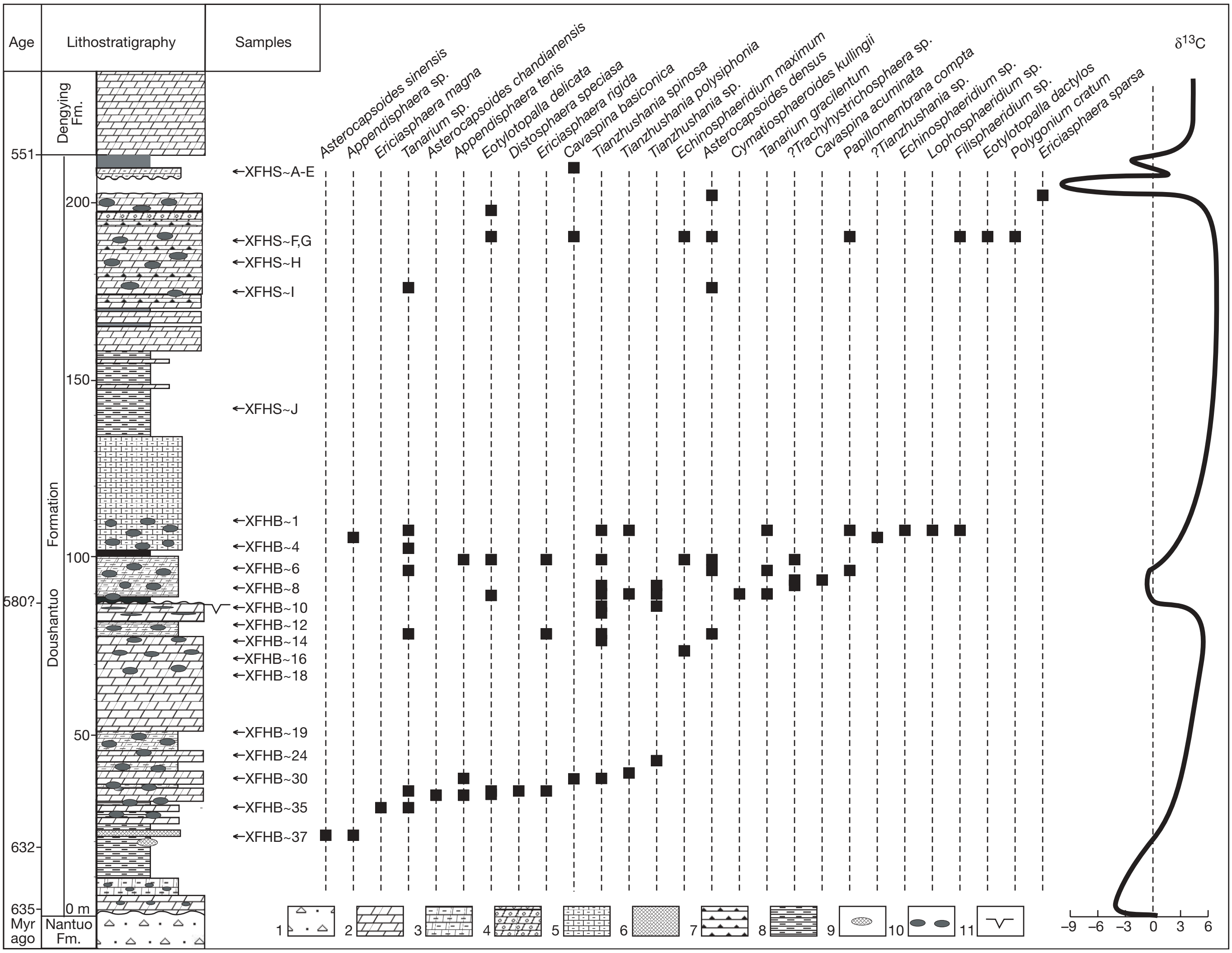


Figure 2 | Xiaofenghe section of the Doushantuo Formation, showing the stratigraphic distribution of large acanthomorph acritarchs. The generalized carbon isotopic curve is based on data published in ref. 20; ages are those reported in ref. 9. 1, diamictite; 2, dolostone; 3, muddy dolostone; 4, oolitic dolostone; 5, muddy limestone; 6, phosphorite; 7, chert; 8, black shale; 9, phosphorite concretions; 10, chert nodules; 11, sequence boundary. Fm., formation.

It has been proposed that the acritarch genus *Tianzhushania* (vesicles of 300–700 μm diameter, each with a wall ornamented by numerous cylindrical processes and surrounded by an outer membrane) represents the external hull of fossil eggs and embryos[17]. This conclusion is based on petrological and scanning electron micrograph observations that link *Tianzhushania* to spheroidal fossils with cerebroid to tuberuclate surfaces that are interpreted as egg envelopes[2]. One *Tianzhushania* vesicle with two cells inside are known, but until now no contents with higher cell numbers have been discovered[17].

New collections from the Xiaofenghe section of the Doushantuo Formation (Fig. 2), along the eastern limb of the Huangling anticline near Yichang, China, confirm and extend previous interpretations[17], showing that cleavage, at least through the 16-cell stage, occurred within *Tianzhushania* hulls (Fig. 1). The 1, 2, 4, 8 and 16-cell stages preserved inside *Tianzhushania* vesicles are identical to other populations interpreted as early cleavage embryos[2]. Thus, the life cycle of the organisms that produced the embryos also included formation of a complex cyst or hull. This is inconsistent with the interpretation of the populations as giant sulphur bacteria; indeed, we know of no bacteria that produce acanthomorphic spores of the size, morphology and preservability exhibited by the Doushantuo structures[12]. The preserved processes observed cannot be bacterial fimbriae because they are hollow and arise from the surface of an encompassing cyst, not from surfaces interpreted as cell walls in ref. 8. In contrast, the combined morphological evidence of ornamented walls and multicellular contents is fully consistent with the interpretation of the Doushantuo fossils as animals and suggests that a resting stage was part of early-animal life cycles. Other large, spinose microfossils in Doushantuo phosphorites also preserve multicellular contents[13], so perhaps a greater record of animal diversity exists among the diverse acanthomorphic cyst walls preserved in Doushantuo and other Ediacaran rocks[12,18].

The recognition that *Tianzhushania* acritarchs and phosphatized embryos detail sections of a single life cycle[19] allows us to investigate more fully the stratigraphic distribution of early animals, even where embryos are not preserved. In the Doushantuo Formation, embryo populations are restricted to beds that lie above a major sequence boundary within the succession[4,13]. Mid-Doushantuo sequence boundaries recorded across the formation's known area of outcrop have been correlated[20] and interpreted in terms of sea level change governed by Gaskiers glaciation in eastern North America and elsewhere[9]. Pb–Pb analyses of Doushantuo phosphorites provide ages of from 599 ± 4 Myr ago[21] to 576 ± 14 Myr ago[22] for upper Doushantuo deposition. If the correlation of mid-Doushantuo sequence boundary with Gaskiers glaciation is correct, however, the diverse acritarchs, algae and embryos in Doushantuo rocks must be younger than 580 Myr old and older than a 551.7 ± 07-Myr-old ash bed near the top of the formation (Fig. 2)[9].

A bed-by-bed census of microfossils throughout the 220 m Xiaofeng succession indicates that *Tianzhushania* acritarchs occur low in the formation, within chert and phosphorite nodules in dolomitic shales only about 30 m above glaciogenic rocks of the underlying Nantuo Formation (Fig. 2; see also ref. 23). U–Pb zircon analyses provide a date for an ash bed within the cap carbonate at the formation's base of 635.2 ± 0.6 Myr ago; a second ash bed 5 m below the earliest *Tianzhushania* fossils is dated as being 632.5 ± 0.5 Myr old[9]. Thus, organisms that produced hulls that are indistinguishable from those containing embryos, higher in the Doushantuo succession, existed within about 3 million years of snowball ice sheet decay[24]. Other taxa of large acanthomorphic microfossils occur still lower in the section, below the 632.5 ± 0.5-Myr-old ash bed (Fig. 2)[23].

In summary, the discovery that large, morphologically complex cyst walls or hulls contain fossils that are interpreted as the early cleavage stages of animal embryos supports interpretation of both embryos and cyst walls as metazoan. The expanded set of search images for early animal fossils afforded by this association provides evidence that at least stem group metazoans were components of earliest Ediacaran ecosystems, later expanding markedly, perhaps in association with rising oxygen levels[25–27].

Received 8 January; accepted 13 February 2007.

1. Xiao, S., Zhang, Y. & Knoll, A. H. Three-dimensional preservation of algae and animal embryos in a Neoproterozoic phosphorite. *Nature* **391**, 553–558 (1998).
2. Xiao, S. & Knoll, A. H. Phosphatized animal embryos from the Neoproterozoic Doushantuo Formation at Weng'an, Guizhou, South China. *J. Paleontol.* **74**, 767–788 (2000).
3. Chen, J. *et al.* Small bilaterian fossils from 40 to 55 million years before the Cambrian. *Science* **305**, 218–222 (2006).
4. Dornbos, S. Q. *et al.* Environmental controls on the taphonomy of phosphatized animals and animal embryos from the Neoproterozoic Doushantuo Formation, southwest China. *Palaios* **21**, 3–14 (2006).
5. Raff, E. C., Villinski, J. T., Turner, F. R., Donoghue, P. C. J. & Raff, R. A. Experimental taphonomy shows the feasibility of fossil embryos. *Proc. Natl Acad. Sci. USA* **103**, 5846–5851 (2006).
6. Hagadorn, J. W. *et al.* Cellular and subcellular structure of Neoproterozoic animal embryos. *Science* **314**, 291–294 (2006).
7. Donoghue, P. C. J. *et al.* Fossilized embryos are widespread but the record is temporally and taxonomically biased. *Evol. Dev.* **8**, 232–238 (2006).
8. Bailey, J. V., Joye, S. B., Kalanetra, K. M., Flood, B. E. & Corsetti, F. A. Evidence of giant sulfur bacteria in Neoproterozoic phosphorites. *Nature* **445**, 198–201 (2007).
9. Condon, D. *et al.* U-Pb ages from the Neoproterozoic Doushantuo Formation, China. *Science* **308**, 95–98 (2005).
10. Donoghue, P. C. J. Embryonic identity crisis. *Nature* **445**, 155–156 (2007).
11. Xiao, S., Hagadorn, J. W., Zhou, C. & Yuan, X. Rare helical spheroidal fossils from the Doushantuo Lagerstätte: Ediacaran animal embryos come of age? *Geology* **35**, 115–118 (2007).
12. Knoll, A. H., Javaux, E. J., Hewitt, D. & Cohen, P. Eukaryotic organisms in Proterozoic oceans. *Phil. Trans. Roy. Soc. B* **361**, 1023–1038 (2006).
13. Zhang, Y., Yin, L., Xiao, S. & Knoll, A. H. Permineralized fossils from the terminal Proterozoic Doushantuo Formation, south China. *J. Paleontol.* **72**, (Suppl. to no. 4), 1–52 (1998).
14. Yuan, X. & Hofmann, H. J. New micropaleontological data from Neoproterozoic Sinica Doushantuo phosphorite rocks, Weng'an, Guizhou Province, southwestern China. *Alcheringa* **22**, 189–222 (1998).
15. Van Waveren, I. & Marcus, N. H. Morphology of recent copepod egg envelopes from Turkey Point, Gulf of Mexico, and their implications for acritarch affinity. *Spec. Pap. Paleontol.* **48**, 111–124 (1993).
16. Buckland-Nicks, J. Hull cupules of chiton eggs: parachute structures and sperm focusing devices? *Biol. Bull.* **184**, 269–276 (1993).
17. Yin, C., Bengtson, S. & Yue, Z. Silicified and phosphatized *Tianzhushania*, spheroidal microfossils of possible animal origin from the Neoproterozoic of South China. *Acta Palaeontol. Pol.* **49**, 1–12 (2004).
18. Grey, K. Ediacaran palynology of Australia. *Mem. Assoc. Australas. Palaeontologists* **31**, 1–439 (2005).
19. Marcus, N. H. & Boero, F. Minireview: The importance of benthic–pelagic coupling and the forgotten role of life cycles in coastal aquatic systems. *Limnol. Oceanogr.* **43**, 763–768 (1998).
20. Zhu, M., Zhang, J. & Yang, A. Integrated Ediacaran (Sinian) chronostratigraphy of South China. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* (in the press).
21. Barfod, G. H. *et al.* New Lu-Hf and Pb-Pb age constraints on the earliest animal fossils. *Earth Planet. Sci. Lett.* **201**, 203–212 (2002).
22. Chen, D., Dong, W. Q., Qi, L., Chen, G. Q. & Chen, X. P. Pb-Pb ages of Neproterozoic Doushantuo phosphorites in South China: constraints on early metazoan evolution and glaciation events. *Precambr. Res.* **132**, 123–222 (2004).
23. Zhou, C., Xie, G., Mcfadden, K., Xiao, S. & Yuan, X. The diversification and extinction of Doushantuo–Pertatataka acritarchs in South China: causes and biostratigraphic significance. *Geol. J.* **41**, 1–34 (2006).
24. Hoffman, P. F., Kaufman, A. J., Halverson, G. P. & Schrag, D. P. A Neoproterozoic snowball earth. *Science* **281**, 1342–1346 (1998).
25. Voreb'eva, N. G., Sergeev, V. N. & Knoll, A. H. Diverse Ediacaran acritarchs from the margin of the East European Platform. *Geology* (submitted).
26. Canfield, D. E., Poulton, S. W. & Narbonne, G. M. Late Neoproterozoic deep-ocean oxygenation and the rise of animal life. *Science* 10.1126/science.1135013 (2006).
27. Fike, D. A., Grotzinger, J. P., Pratt, L. M. & Summons, R. E. Oxidation of the Ediacaran ocean. *Nature* **444**, 744–747 (2006).

Acknowledgements This work was supported by the Chinese Academy of Sciences, the National Natural Science Foundation of China, the Major Basic Research Projects of MST of China, and an NSF Grant. We thank S. Xiao and P. Cohen for discussions.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to L.Y. (leimingyin@yahoo.com.cn).

Vol 446 | 5 April 2007 | doi:10.1038/nature05670

LETTERS

Quantifying social group evolution

Gergely Palla[1], Albert-László Barabási[2] & Tamás Vicsek[1,3]

The rich set of interactions between individuals in society[1–7] results in complex community structure, capturing highly connected circles of friends, families or professional cliques in a social network[3,7–10]. Thanks to frequent changes in the activity and communication patterns of individuals, the associated social and communication network is subject to constant evolution[7,11–16]. Our knowledge of the mechanisms governing the underlying community dynamics is limited, but is essential for a deeper understanding of the development and self-optimization of society as a whole[17–22]. We have developed an algorithm based on clique percolation[23,24] that allows us to investigate the time dependence of overlapping communities on a large scale, and thus uncover basic relationships characterizing community evolution. Our focus is on networks capturing the collaboration between scientists and the calls between mobile phone users. We find that large groups persist for longer if they are capable of dynamically altering their membership, suggesting that an ability to change the group composition results in better adaptability. The behaviour of small groups displays the opposite tendency—the condition for stability is that their composition remains unchanged. We also show that knowledge of the time commitment of members to a given community can be used for estimating the community's lifetime. These findings offer insight into the fundamental differences between the dynamics of small groups and large institutions.

The data sets we consider are (1) the monthly list of articles in the Cornell University Library e-print condensed matter (cond-mat) archive spanning 142 months, with over 30,000 authors[25], and (2) the record of phone calls between the customers of a mobile phone company spanning 52 weeks (accumulated over two-week-long periods), and containing the communication patterns of over 4 million users. Both types of collaboration events (a new article or a phone call) document the presence of social interaction between the involved individuals (nodes), and can be represented as (time-dependent) links. The extraction of the changing link weights from the primary data is described in Supplementary Information. In Fig. 1a, b we show the local structure at a given time step in the two networks in the vicinity of a randomly chosen individual (marked by a red frame). The communities (social groups represented by more densely interconnected parts within a network of social links) are colour coded, so that black nodes/edges do not belong to any community, and those that simultaneously belong to two or more communities are shown in red.

The two networks have rather different local structure: the collaboration network of scientists emerges as a one-mode projection of the bipartite graph between authors and papers, so it is quite dense and the overlap between communities is very significant. In contrast, in the phone-call network the communities are less interconnected and are often separated by one or more inter-community nodes/edges. Indeed, whereas the phone record captures the communication between two people, the publication record assigns to all individuals that contribute to a paper a fully connected clique. As a result, the phone data are dominated by single links, whereas the co-authorship data have many dense, highly connected neighbourhoods. Furthermore, the links in the phone network correspond to instant communication events, capturing a relationship as it happens. In contrast, the co-authorship data

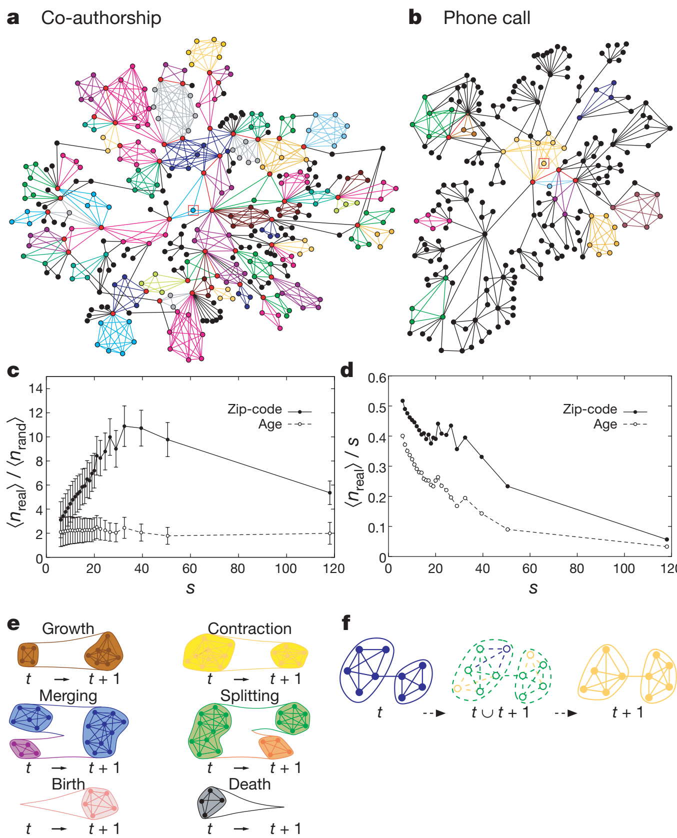


Figure 1 | Structure and schematic dynamics of the two networks considered. a, The co-authorship network. The figure shows the local community structure at a given time step in the vicinity of a randomly selected node. **b**, As **a** but for the phone-call network. **c**, The filled black symbols correspond to the average size of the largest subset of members with the same zip-code, $\langle n_{\rm real}\rangle$, in the phone-call communities divided by the same quantity found in random sets, $\langle n_{\rm rand}\rangle$, as a function of the community size, s. Similarly, the open symbols show the average size of the largest subset of community members with an age falling in a three-year time window, divided by the same quantity in random sets. The error bars in both cases correspond to $\langle n_{\rm real}\rangle/(\langle n_{\rm rand}\rangle + \sigma_{\rm rand})$ and $\langle n_{\rm real}\rangle/(\langle n_{\rm rand}\rangle - \sigma_{\rm rand})$, where $\sigma_{\rm rand}$ is the standard deviation in the case of the random sets. **d**, The $\langle n_{\rm real}\rangle/s$ as a function of s, for both the zip-code (filled black symbols) and the age (open symbols). **e**, Possible events in community evolution. **f**, The identification of evolving communities. The links at t (blue) and the links at $t+1$ (yellow) are merged into a joint graph (green). Any CPM community at t or $t+1$ is part of a CPM community in the joined graph, so these can be used to match the two sets of communities.

[1]Statistical and Biological Physics Research Group of the HAS, Pázmány P. stny. 1A, H-1117 Budapest, Hungary. [2]Center for Complex Network Research and Departments of Physics and Computer Science, University of Notre Dame, Indiana 46566, USA. [3]Department of Biological Physics, Eötvös University, Pázmány P. stny. 1A, H-1117 Budapest, Hungary.

record the results of a long-term collaboration process. These fundamental differences suggest that any common features of the community evolution in the two networks represent potentially generic characteristics of community formation, rather than being rooted in the details of the network representation or data collection process.

The communities at each time step were extracted using the clique percolation method[23,24] (CPM). The key features of the communities obtained by the CPM are that their members can be reached through well connected subsets of nodes, and that the communities may overlap (share nodes with each other). This latter property is essential, as most networks are characterized by overlapping and nested communities[6,23]. As a first step, it is important to check if the uncovered communities correspond to groups of individuals with a shared common activity pattern. For this purpose, we compared the average weight of the links inside communities, w_c, to the average weight of the inter-community links, w_{ic}. For the co-authorship network w_c/w_{ic} is about 2.9, whereas for the phone-call network the difference is even more significant, as $w_c/w_{ic} \approx 5.9$, indicating that the intensity of collaboration/communication within a group is significantly higher than with contacts belonging to a different group[26–28]. Although for co-authors the quality of the clustering can be directly tested by studying their publication records in more detail, in the phone-call network personal information is not available. In this case the zip-code and the age of the users provide additional information for checking the homogeneity of the communities. According to Fig. 1c, the $\langle n_{real}\rangle/\langle n_{rand}\rangle$ ratio is significantly larger than 1 for both the zip-code and the age, indicating that communities have a tendency to contain people from the same generation and living in the same neighbourhood ($\langle n_{real}\rangle$ is the size of the largest subset of people having the same zip-code averaged over time steps and the set of available communities, while $\langle n_{rand}\rangle$ represents the same average but with randomly selected users). It is of specific interest that $\langle n_{real}\rangle/\langle n_{rand}\rangle$ for the zip-code has a prominent peak at community size $s \approx 35$, suggesting that communities of this size are geographically the most homogeneous ones. However, as Fig. 1d shows, the situation is more complex: on average, the smaller communities are more homogeneous in respect of both the zip-code and the age, but there is still a noticeable peak at $s \approx 30$–35 for the zip-code. In summary, the phone-call communities uncovered by the CPM tend to contain individuals living in the same neighbourhood, and having a comparable age, a homogeneity that supports the validity of the uncovered community structure. Further support is given in Supplementary Information.

The basic events that may occur in the life of a community are shown in Fig. 1e: a community can grow or contract; groups may merge or split; new communities are born while others may disappear. We have developed a method for the appropriate matching (between the subsequent states of the evolving communities) from the information available for relatively distant points in time only (see Methods).

After determining the dynamically changing community structure, we first consider two basic quantities characterizing a community: its size s and its age τ, representing the time passed since its birth. The quantities s and τ are positively correlated: larger communities are on average older (Fig. 2a). Next we used the auto-correlation function, $C(t)$, to quantify the relative overlap between two states of the same community $A(t)$ at t time steps apart:

$$C(t) \equiv \frac{|A(t_0) \cap A(t_0+t)|}{|A(t_0) \cup A(t_0+t)|} \qquad (1)$$

where $|A(t_0) \cap A(t_0+t)|$ is the number of common nodes (members) in $A(t_0)$ and $A(t_0+t)$, and $|A(t_0) \cup A(t_0+t)|$ is the number of nodes in the union of $A(t_0)$ and $A(t_0+t)$. Figure 2b shows the average time dependent auto-correlation function for communities born with different sizes. The data indicate that the collaboration network is more 'dynamic' ($\langle C(t)\rangle$ decays faster). We also find that in both networks, the auto-correlation function decays faster for the larger communities, showing that the membership of the larger communities is changing at a higher rate. In contrast, small communities change at a smaller rate, their composition being more or less static. To quantify this aspect of community evolution, we define the stationarity ζ of a community as the average correlation between subsequent states:

$$\zeta \equiv \frac{\sum_{t=t_0}^{t_{max}-1} C(t,t+1)}{t_{max}-t_0-1} \qquad (2)$$

where t_0 denotes the birth of the community, and t_{max} is the last step before the extinction of the community. Thus, $1-\zeta$ represents the average ratio of members changed in one step.

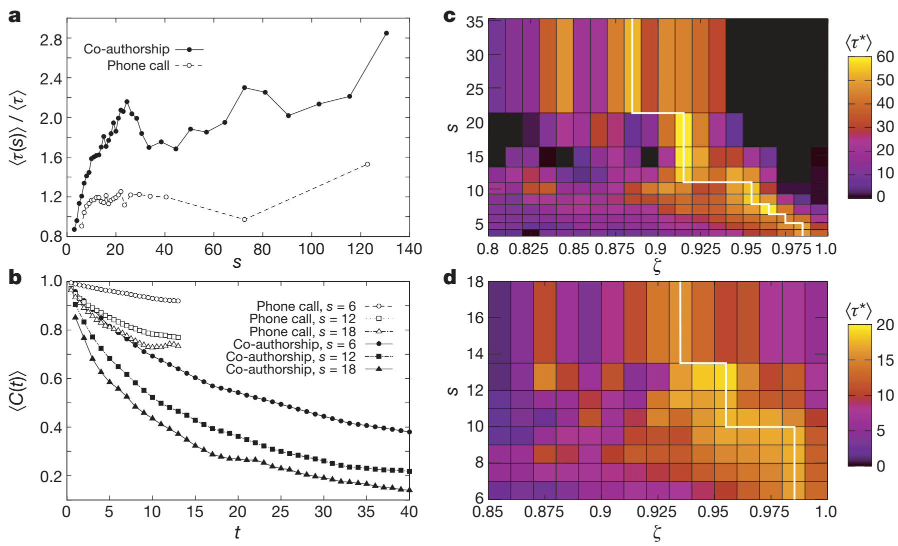


Figure 2 | Characteristic features of community evolution. **a**, The age τ of communities with a given size (number of people) s, averaged over the set of available communities and the time steps, divided by the average age of all communities, $\langle\tau\rangle$, as a function of s. The increasing nature of the plot indicates that larger communities are on average older. **b**, The auto-correlation function $C(t)$ of communities with different sizes averaged over the communities and t_0. The unit of time, t, is two weeks, thus, for the co-authorship network, where the data samples were taken monthly, the $C(t)$ values are shown for every other time step. **c**, The life-span τ^* averaged over the communities as a function of the stationarity ζ and the community size s for the co-authorship network. (The communities still living at the last available time step in the data set were excluded from this investigation.) The peak in $\langle\tau^*\rangle$ is close to $\zeta = 1$ for small sizes, whereas it is shifted towards lower ζ values for large sizes. **d**, Similar results found in the phone-call network. In **c** and **d**, the white line corresponds to the optimal stationarity.

We observe an interesting effect when investigating the relationship between the lifetime τ^* (the number of steps between the birth and the disintegration of a community), the stationarity and the community size. The lifetime can be viewed as a simple measure of 'fitness': communities having higher fitness have an extended life, whereas the ones with small fitness quickly disintegrate. In Fig. 2c, d we show the average life-span $\langle\tau^*\rangle$ (colour coded) as a function of the stationarity ζ and the community size s (both s and ζ were binned). In both networks, for small community sizes the highest average life-span is at a stationarity value very close to one, indicating that for small communities it is optimal to have static, time-independent membership. On the other hand, the peak in $\langle\tau^*\rangle$ is shifted towards low ζ values for large communities, suggesting that for these the optimal regime is to be dynamic, that is, to have a continually changing membership.

To illustrate the difference in the optimal behaviour (a pattern of membership dynamics leading to extended lifetime) of small and large communities, in Fig. 3 we show the time evolution of four communities from the co-authorship network. As Fig. 3 indicates, a typical small and stationary community undergoes minor changes, but lives for a long time. This is well illustrated by the snapshots of the community structure, showing that the community's stability is conferred by a core of three individuals representing a collaborative group spanning over 52 months. In contrast, a small community with high turnover of its members has a lifetime of nine time steps only (Fig. 3b). The opposite is seen for large communities: a large stationary community disintegrates after four time steps (Fig. 3c). In contrast, a large non-stationary community whose members change dynamically, resulting in significant fluctuations in both size and composition, has a quite extended lifetime (Fig. 3d).

The different stability rules followed by the small and large communities raise an important question: could the inspection of

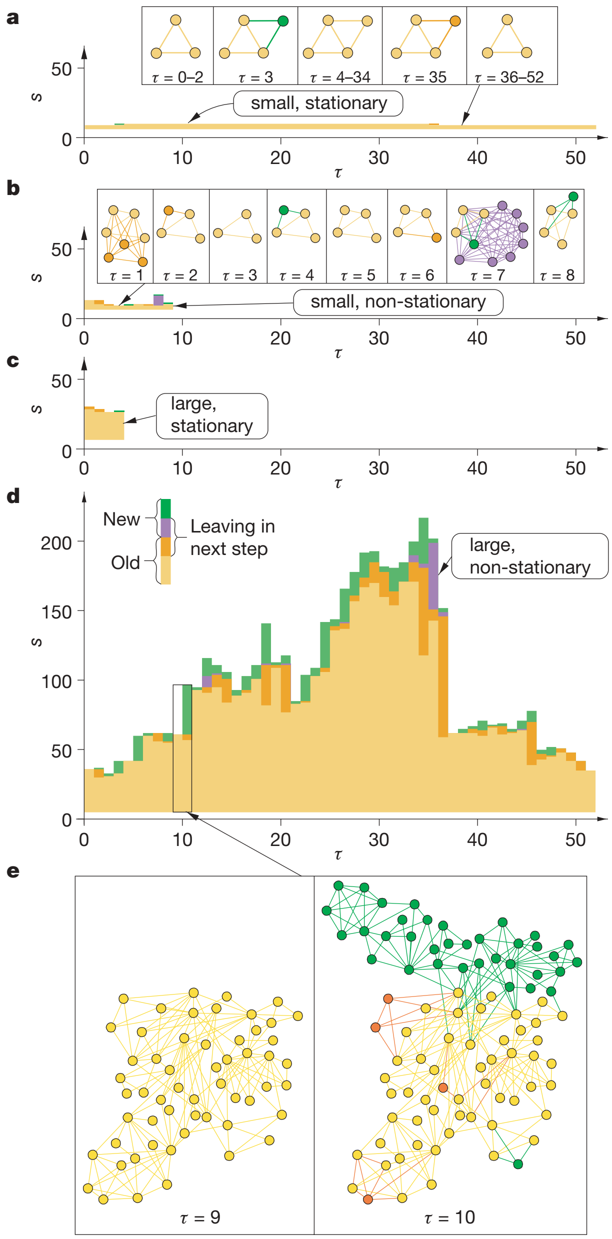


Figure 3 | Evolution of four types of communities in the co-authorship network. The height of the columns corresponds to the actual community size, s, and within one column the yellow colour indicates the number of 'old' nodes (that have been present in the community at least in the previous time step as well), while newcomers are shown with green. The members abandoning the community in the next time step are shown with orange or purple colour, depending on whether they are old or new. (This latter type of member joins the community for only one time step.) We show a small and stationary community (**a**), a small and non-stationary community (**b**), a large and stationary community (**c**) and, finally, a large and non-stationary community (**d**). A mainly growing stage (two time steps) in the evolution of the last community is detailed in **e**.

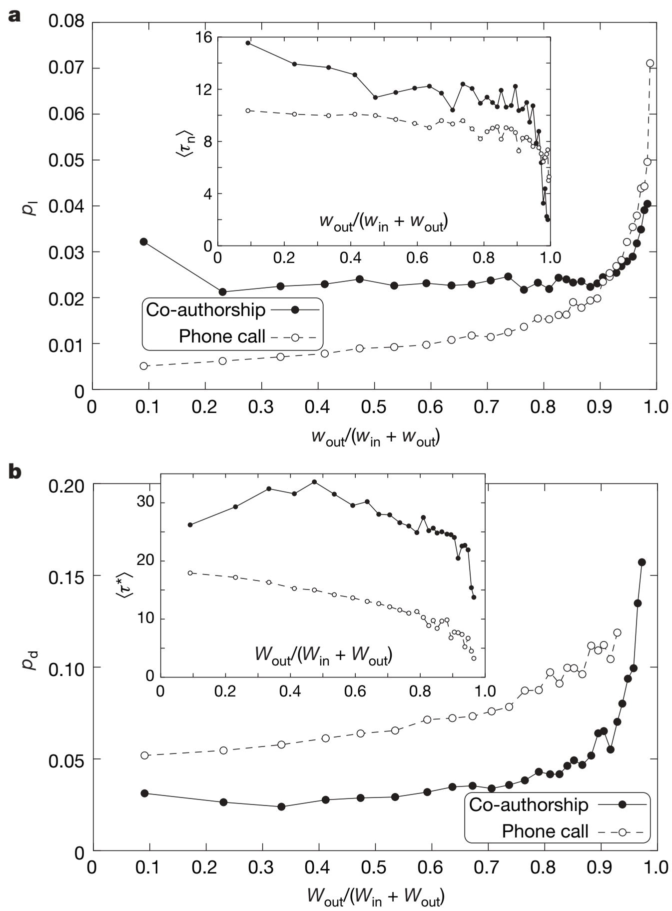


Figure 4 | Effects of links between communities. **a**, The probability p_l of a member abandoning the community in the next step as a function of the ratio of the member's aggregated link weights to other parts of the network (w_{out}) and the member's total aggregated link weight ($w_{in} + w_{out}$). The inset shows the average time spent in the community by the nodes, $\langle\tau_n\rangle$, as a function of $w_{out}/(w_{in} + w_{out})$. **b**, The probability p_d of a community disintegrating in the next step as a function of the ratio of the aggregated weights of links from the community to other parts of the network (W_{out}) and the aggregated weights of all links starting from the community ($W_{in} + W_{out}$). The inset shows the average life-time $\langle\tau^*\rangle$ of communities as a function of $W_{out}/(W_{in} + W_{out})$.

a community itself predict its future? To address this issue, for each member in a community we measured the total weight of this member's connections (a measure of how much a member is committed) to outside of the community (w_{out}) as well as to members belonging to the same community (w_{in}). We then calculated the probability that the member will abandon the community as a function of the $w_{out}/(w_{in} + w_{out})$ ratio. As Fig. 4a shows for both networks, if the relative commitment of a user to individuals outside a given community is higher, then it is more likely that he/she will leave the community. In parallel, the average time spent in the community by the nodes, $\langle\tau_n\rangle$, is a decreasing function of the above ratio (Fig. 4a inset). As Fig. 4a shows, those with the least commitment have a quickly growing likelihood of leaving the community. Taking this idea from individuals to communities, we measured for each community the total weight of links from the members to others, outside of the community (W_{out}), as well as the aggregated link weight inside the community (W_{in}). We find that the probability for a community to disintegrate in the next step increases as a function of $W_{out}/(W_{in} + W_{out})$ (Fig. 4b) and the lifetime of a community decreases for large $W_{out}/(W_{in} + W_{out})$ ratios (Fig. 4b inset). However, an interesting observation is that, whereas the lifetime of the phone-call communities for moderate levels is relatively insensitive to outside commitments, the lifetime of the collaboration communities possesses a maximum at intermediate levels of inter-collaborations (collaboration between colleagues who belong to different communities). These results suggest that a tracking of the individual's as well as the community's relative commitment to the other members of the community provides a clue for predicting the community's fate.

In summary, our results indicate the significant difference between smaller collaborative or friendship circles and institutions. At the heart of small cliques are a few strong relationships, and as long as these persist, the community around them is stable. It appears to be almost impossible to maintain this strategy for large communities. Our calculations show that the condition for stability of large communities is continuous change, so that after some time practically all members are exchanged. Such loose, rapidly changing communities are reminiscent of institutions, which can continue to exist even after all members have been replaced by new members. For example, in a few years most members of a school or a company could change, yet the school and the company will be detectable as a distinct community at any time during its existence.

METHODS

Locating communities. In the CPM method, a community is defined as a union of all k-cliques (complete sub-graphs of size k) that can be reached from each other through a series of adjacent k-cliques (where adjacency means sharing $k-1$ nodes)[24,29]. When applied to weighted networks, the CPM has two parameters: the k-clique size k (in Fig. 1a, b we show the communities for $k = 4$) and the weight threshold w^* (links weaker than w^* are ignored). The criterion for selecting these parameters is discussed in the Supplementary Information.

Identifying evolving communities. The basic idea of the algorithm developed by us to identify community evolution is shown in Fig. 1f. For each pair of consecutive time steps t and $t+1$, we construct a joint graph consisting of the union of links from the corresponding two networks, and extract the CPM community structure of this joint network (I. Derényi, personal communication). Any community from either the t or the $t+1$ snap-shot is contained in exactly one community in the joint graph, as by adding links to a network, the CPM communities can only grow, merge or remain unchanged. Thus, the communities in the joint graph provide a natural connection between the communities at t and at $t+1$. If a community in the joint graph contains a single community from t and a single community from $t+1$, then they are matched. If the joint group contains more than one community from either time steps, the communities are matched in descending order of their relative node overlap (see Supplementary Information).

Received 2 January; accepted 8 February 2007.

1. Watts, D. J. & Strogatz, S. H. Collective dynamics of 'small-world' networks. *Nature* **393,** 440–442 (1998).
2. Barabási, A.-L. & Albert, R. Emergence of scaling in random networks. *Science* **286,** 509–512 (1999).
3. Scott, J. *Social Network Analysis: A Handbook* 2nd edn (Sage Publications, London, 2000).
4. Dorogovtsev, S. N. & Mendes, J. F. F. *Evolution of Networks: From Biological Nets to the Internet and WWW* (Oxford Univ. Press, Oxford, 2003).
5. Watts, D. J., Dodds, P. S. & Newman, M. E. J. Identity and search in social networks. *Science* **296,** 1302–1305 (2002).
6. Faust, K. in *Models and Methods in Social Network Analysis* (eds Carrington, P., Scott, J. & Wasserman, S.) Ch. 7 (Cambridge Univ. Press, New York, 2005).
7. Liljeros, F., Edling, Ch. R., Amaral, L. A. N., Stanley, H. E. & Aberg, Y. The web of human sexual contacts. *Nature* **411,** 907–908 (2001).
8. Shiffrin, R. M. & Börner, K. Mapping knowledge domains. *Proc. Natl Acad. Sci. USA* **101** (suppl. 1), 5183–5185 (2004).
9. Newman, M. E. J. Detecting community structure in networks. *Eur. Phys. J. B* **38,** 321–330 (2004).
10. Radicchi, F., Castellano, C., Cecconi, F., Loreto, V. & Parisi, D. Defining and identifying communities in networks. *Proc. Natl Acad. Sci. USA* **101,** 2658–2663 (2004).
11. Barabási, A.-L. *et al.* Evolution of the social network of scientific collaborations. *Physica A* **311,** 590–614 (2002).
12. Holme, P., Edling, Ch. R. & Liljeros, F. Structure and time-evolution of an internet dating community. *Soc. Networks* **26,** 155–174 (2004).
13. Ebel, H., Davidsen, J. & Bornholdt, S. Dynamics of social networks. *Complexity* **8,** 24–27 (2002).
14. Wagner, C. S. & Leydesdorff, L. Network structure, self-organization, and the growth of international collaboration in science. *Res. Policy* **34,** 1608–1618 (2005).
15. Yeung, Y.-Y., Liu, T. C.-Y. & Ng, P.-H. A social network analysis of research collaboration in physics education. *Am. J. Phys.* **73,** 145–150 (2005).
16. Newman, M. E. J. & Park, J. Why social networks are different from other types of networks. *Phys. Rev. E* **68,** 036122 (2003).
17. Guimerá, R., Danon, L., Diaz-Guilera, A., Giralt, F. & Arenas, A. Self-similar community structure in organisations. *Phys. Rev. E* **68,** 065103 (2003).
18. Hopcroft, J., Khan, O., Kulis, B. & Selman, B. Tracking evolving communities in large linked networks. *Proc. Natl Acad. Sci. USA* **101,** 5249–5253 (2004).
19. Guimerá, R., Uzzi, B., Spiro, J. & Amaral, L. A. N. Team assembly mechanisms determine collaboration network structure and team performance. *Science* **308,** 697–702 (2005).
20. Li, Ch. & Maini, Ph. K. An evolving network model with community structure. *J. Phys. Math. Gen.* **38,** 9741–9749 (2005).
21. Pollner, P., Palla, G. & Vicsek, T. Preferential attachment of communities: The same principle, but a higher level. *Europhys. Lett.* **73,** 478–484 (2006).
22. Kossinets, G. & Watts, D. J. Empirical analysis of an evolving social network. *Science* **311,** 88–90 (2006).
23. Palla, G., Derényi, I., Farkas, I. & Vicsek, T. Uncovering the overlapping community structure of complex networks in nature and society. *Nature* **435,** 814–818 (2005).
24. Derényi, I., Palla, G. & Vicsek, T. Clique percolation in random networks. *Phys. Rev. Lett.* **94,** 160202 (2005).
25. Warner, S. E-prints and the open archives initiative. *Library Hi Tech* **21,** 151–158 (2003).
26. Granovetter, M. S. The strength of weak ties. *Am. J. Sociol.* **78,** 1360–1380 (1973).
27. Csermely, P. *Weak Links* (Springer, Heidelberg, 2006).
28. Onnela, J.-P. *et al.* Structure and tie strengths in mobile communication networks. Preprint at ⟨http://lanl.arxiv.org/abs/physics/0610104⟩ (2006).
29. Everett, M. G. & Borgatti, S. P. Analyzing clique overlap. *Connections* **21,** 49–61 (1998).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank I. Derényi for suggestions, and G. Szabó and I. Farkas for their assistance with the primary phone-call and co-authorship data sets, respectively. G.P and T.V are supported by OTKA; A.-L.B. is supported by the James S. McDonnell Foundation and the National Science Foundation within the DDDAS and ITR programmes.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Correspondence and requests for materials should be addressed to T.V. (vicsek@angel.elte.hu).

nature Vol 446|5 April 2007|doi:10.1038/nature05685

LETTERS

Antibiotic interactions that select against resistance

Remy Chait[1], Allison Craney[1] & Roy Kishony[1,2]

Multidrug combinations are increasingly important in combating the spread of antibiotic-resistance in bacterial pathogens[1–3]. On a broader scale, such combinations are also important in understanding microbial ecology and evolution[4,5]. Although the effects of multidrug combinations on bacterial growth have been studied extensively, relatively little is known about their impact on the differential selection between sensitive and resistant bacterial populations[1,6,7]. Normally, the presence of a drug confers an advantage on its resistant mutants in competition with the sensitive wild-type population[1]. Here we show, by using a direct competition assay between doxycycline-resistant and doxycycline-sensitive *Escherichia coli*, that this differential selection can be inverted in a hyper-antagonistic class of drug combinations. Used in such a combination, a drug can render the combined treatment selective against the drug's own resistance allele. Further, this inversion of selection seems largely insensitive to the underlying resistance mechanism and occurs, at sublethal concentrations, while maintaining inhibition of the wild type. These seemingly paradoxical results can be rationalized in terms of a simple geometric argument. Our findings demonstrate a previously unappreciated feature of the fitness landscape for the evolution of resistance and point to a trade-off between the effect of drug interactions on absolute potency and the relative competitive selection that they impose on emerging resistant populations.

The rapid evolution of bacterial drug resistance and the alarming slowdown in development of new antibiotics is spurring attention towards multidrug treatments[2,8]. Drug combinations are classified as synergistic, additive or antagonistic, according to whether the combined effect of the drugs is larger than, equal to or smaller than the effect predicted by their individual activities (Fig. 1)[8,9]. In some cases the effect of the drug combination is even less than that of one of the drugs by itself; we refer to such hyper-antagonistic interactions as suppression (Fig. 1)[10,11]. Normally, we expect resistance to even one of the drugs in a multidrug treatment to confer an advantage on the bacteria. However, an intriguing hypothesis is that in suppressive multidrug treatments, resistance to one of the drugs could actually have the opposite effect. In such cases, although resistance would indeed diminish the burden imposed by one of the drugs, it may also remove the suppression, rendering the combined treatment more effective against the resistant mutant than against the wild type.

Motivated by this hypothesis, we examined the selective pressure imposed on a drug-resistance allele under synergistic versus suppressive drug combinations. As a model, we focused on resistance to doxycycline, a medically important tetracycline antibiotic broadly used to treat a variety of Gram-negative and Gram-positive infections[12]. Doxycycline inhibits protein synthesis by blocking aminoacyl-tRNA binding at the A-site in the 30S ribosomal subunit[13]. Resistance to tetracyclines is typically associated with mobile genetic elements and falls into three classes: efflux pump, ribosomal protection and enzymatic degradation[12,13].

To provide a simple model of the competition between resistant and sensitive strains, we constructed a pair of *Escherichia coli* strains differing only in the chromosomal presence or absence of the Tn10 transposon encoding a well-studied tetracyclines efflux pump (Methods)[14–16]. Although treatment with doxycycline alone obviously confers strong selection for the resistant strain over the sensitive strain, we examined how combinations of doxycycline with other drugs affect this selection pressure. We chose erythromycin and ciprofloxacin, representing the macrolides and quinolones[13] and showing respectively synergistic and suppressive (or antagonistic at low concentrations) interactions with doxycycline[10].

We first tested the effect of the two chosen drug pairs on the growth rate of the doxycycline-sensitive (wild-type) strain. We used a bioluminescence-based assay, which accurately measures bacterial growth rates with a detection sensitivity exceeding that of standard optical density techniques by three orders of magnitude (Methods,

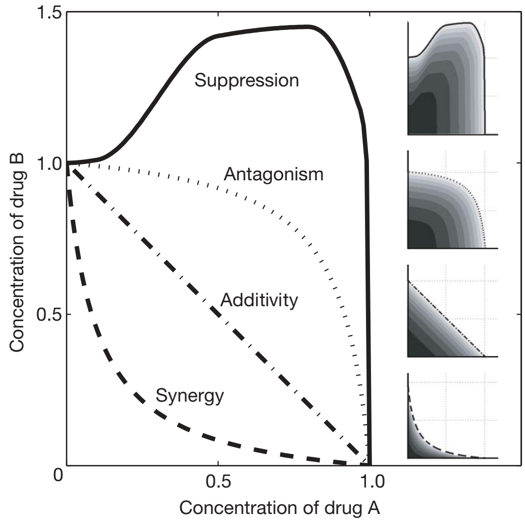


Figure 1 | Schematic representation of synergistic, additive, antagonistic and suppressive drug pairs. Lines of equal effect of the drug combination on growth rate (isoboles) are shown in the two-dimensional concentration space of the two drugs. Loewe additivity[9] assumes that two drugs do not interact if their combined outcome is that expected from a linear interpolation of their two individual outcomes (dash-dotted line). Synergistic drug pairs have a stronger than additive effect corresponding to an isobole below the additive line (dashed line). Antagonistic drug pairs have a less than additive effect (dotted). Suppression interactions (solid line) are a subclass of antagonism in which the combined treatment effect is weaker than that of at least one of the drugs alone. The insets show schematic growth rate maps (darker indicates faster growth) for each interaction type (top to bottom: suppression, antagonism, additivity, synergy). The axes of the insets are identical to those of the main figure. Growth responses to single drugs alone lie along each axis. The isoboles (black lines) shown here indicate drug pair concentrations required to just halt growth.

[1]Department of Systems Biology, Harvard Medical School, 200 Longwood Avenue, Boston, Massachusetts 02115, USA. [2]School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts 02138, USA.

and Fig. 2e)[17,18]. The response maps of the measured growth rates and the minimum inhibitory concentration (MIC) line as a function of the two-dimensional drug concentration show the strong synergy between doxycycline and erythromycin (Fig. 2a, b, green). In contrast, strong suppression is seen for doxycycline and ciprofloxacin, for which, over a range of concentrations, doxycycline relieves the inhibitory effect of ciprofloxacin (Fig. 2c, d, green).

We then repeated the measurement for the doxycycline-resistant mutant (Fig. 2a–d, red). The results indicate a roughly 100-fold increase in MIC for doxycycline. No significant changes in ciprofloxacin or erythromycin MICs and no detectable difference in growth rate in drug-free medium were observed[19,20] (Supplementary Fig. S1). Response maps for the resistant strain under both drug combinations are very similar to the corresponding maps for the wild type, except for a large (about 100-fold) rescaling along the doxycycline concentration axis. This rescaling, reflecting the increased concentration of a drug required for the same level of inhibition, is expected for resistance mechanisms that specifically decrease the intracellular concentration or activity of one of the drugs. Indeed, similar rescaling is observed for doxycycline resistances that are based on enzymatic degradation[12,21] and ribosomal protection[12,22], although it can break down at high drug concentrations (Supplementary Fig. S2).

Whereas effective rescaling of doxycycline concentrations appears under both the synergistic (doxycycline–erythromycin) and the suppressive (doxycycline–ciprofloxacin) drug pairs, its effects on selection for resistance are profoundly different (Fig. 2a, c). In the synergistic case, rescaling of the MIC line along a single axis causes the mutant growth regime to be completely inclusive of the wild type (Fig. 2a); that is, there is no combination of the drugs with which the wild type survives whereas the resistant bacteria perish. In the suppression case, however, rescaling generates just such a region of drug concentrations, at which the wild type grows but the resistant mutant does not (area marked with an asterisk in Fig. 2c; see also schematic illustration in Supplementary Fig. S3a). Similarly, resistances based on enzymatic degradation or ribosomal protection also yield drug-concentration regions in which the sensitive strain grows but the resistant mutants do not (Supplementary Fig. S2). These observations suggest that, in a competitive situation, resistant strains experience positive selection in all regions of the synergistic treatment but may actually be selected against with certain suppressive drug combinations.

To query this differential selection for resistance directly, we set up a competition assay between the sensitive and resistant strains. We used the method developed in ref. 23 to measure the change in ratio of the wild-type and resistant populations by differential labelling with cyan (CFP) and yellow (YFP) fluorescent proteins (Methods). The competition was performed in a 12×16 array of mixed concentrations of the two drugs (Fig. 3). We observed a neutral or very minor fitness cost of doxycycline resistance in drug-free medium, and in ciprofloxacin or erythromycin alone (yellow along the y axis in Fig. 3a, b; see also Supplementary Fig. S1 and Supplementary Tables 1 and 2). As expected, the doxycycline-resistant mutant outcompeted the wild type under doxycycline treatment alone and when doxycycline was synergistically paired with erythromycin (Fig. 3a). Indeed, selection for the resistance allele was even stronger in the synergistic multidrug treatment than under doxycycline alone (for example, adding 50 μg ml^{-1} erythromycin to 0.1 μg ml^{-1} doxycycline increased selection for doxycycline resistance). A possible intuitive explanation is that resistance to doxycycline effectively reduces

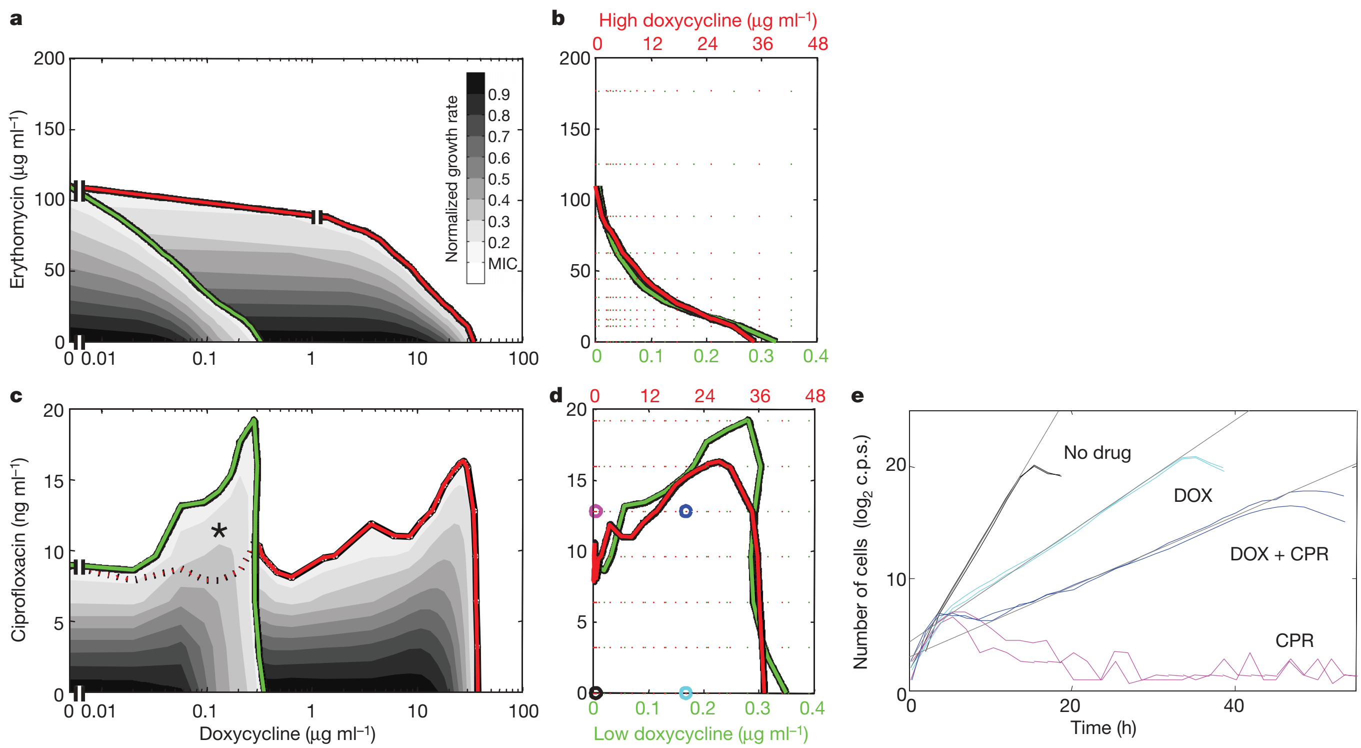


Figure 2 | Rescaling of effective drug concentrations by resistance generates a region exclusive to growth of sensitive bacteria in a suppressive drug combination. **a**–**d**, MIC lines (green, sensitive strain Wyl; red, doxycycline-resistant strain t17yl) and higher growth rate isoboles (grey scale) in synergistic (doxycycline–erythromycin (**a**, **b**)) and suppressive (doxycycline–ciprofloxacin (**c**, **d**)) drug combinations. The MIC lines of the sensitive and resistant strains are similar, except for a rescaling in doxycycline concentrations (linear scales in **b** and **d**). In the synergistic case (**a**), this scaling leaves the growth region of the sensitive strain fully enclosed by that of the resistant strain. In contrast, in the suppressive case, the scaling generates a region in which only the sensitive strain grows (**c**, asterisk). Growth rates were measured at an array of drug concentrations indicated in **b** and **d** (green points, sensitive; red points, resistant) by c.p.s. of bacterial luminescence versus time (see Methods). **e**, Sample growth curves of the sensitive strain at four conditions indicated in **d**: no drug (black), doxycycline only (cyan), ciprofloxacin only (magenta) and the combination (blue). Two replicates and their linear fit (grey lines) are shown.

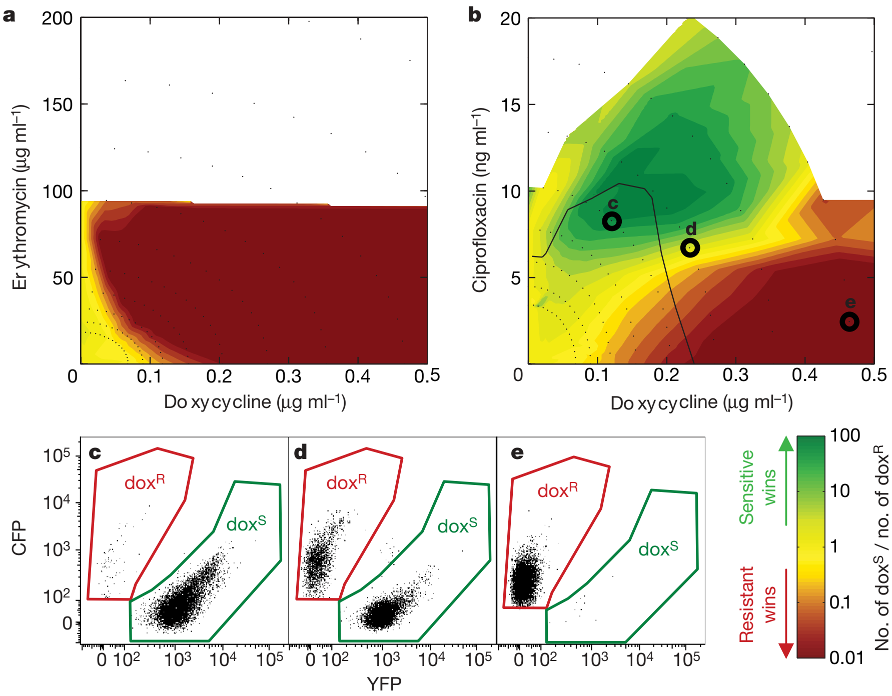


Figure 3 | Competitive selection against resistance in a suppressive drug combination. Doxycycline-sensitive (dox^S) and doxycycline-resistant (dox^R) strains, differentially tagged with CFP and YFP, were inoculated at a 1:1 ratio into an array of drug combinations (**a**, **b**, black dots). Final ratios, reflecting fitness differences (growth and death) between the strains, measured by FACS after 24 h, are shown for combination treatments of doxycycline–erythromycin (synergy, **a**) and doxycycline–ciprofloxacin (suppression, **b**) as indicated in the colour scale at the bottom right. Blank regions indicate no growth. Along a line of constant wild-type inhibition as measured in Fig. 2 (for example 70% inhibition, solid black line in **b**), increasing the doxycycline concentration can select against the resistant mutant even while maintaining inhibition of the wild type. **c–e**, Sample FACS data at points indicated in **b**. The data are consistent across experimental replicates and CFP/YFP marker swaps (data not shown).

not only the doxycycline burden but also the additional burden imposed by the erythromycin–doxycycline synergy. Conversely, under the suppressive drug combination (doxycycline–ciprofloxacin), we found diminished selection for the resistant mutant in comparison with doxycycline alone. Furthermore, there was a clear region of concentrations of ciprofloxacin and doxycycline at which the drug combination actually selected against the resistant genotype (Fig. 3b, green region). Note that in the presence of ciprofloxacin, adding doxycycline generates selection against its own resistant mutant (for example, adding 0.1 µg ml^{-1} doxycycline to 7.5 ng ml^{-1} ciprofloxacin). This relative selection against resistance could be achieved without reducing the absolute level of inhibition of the wild type. At constant wild-type inhibition (Fig. 3b, solid line), the suppressive combination exhibited substantial relative selection against resistance over a broad range of concentrations, whereas ciprofloxacin alone had a neutral or very small effect. These seemingly counterintuitive findings can be explained by our simple geometrical scaling model: when suppression curves such as in Fig. 1 are scaled along the horizontal axis by resistance, they necessarily generate a region of disadvantage for the resistant strain (Supplementary Fig. S3a). In contrast with the phenomenon of resistance–antagonism (specific mutations generate resistance to one drug coupled to hypersusceptibility to another)[24,25], this selection inversion works on uncoupled resistances.

Our data show that in suppressing drug combinations, a drug can be used to exert competitive selection against its own resistance allele. In contrast, synergistic interactions, while increasing absolute potency against both sensitive and resistant strains, also increase relative selection in favour of resistance. These findings point to an inherent tradeoff, where antagonistic combinations, which require a higher dosage and have therefore typically been shunned in clinical therapy[2,3], may have the benefit of reducing and even inverting selection for resistance. Although the molecular mechanisms underlying drug interactions may be complex[3,26,27], suppression between antibiotics is not uncommon[10] (Supplementary Fig. S4). Our simple geometrical approximation anticipates a region of competitive selection against resistance in such suppressive drug combinations when the targeted resistance mechanism works specifically (uniaxially) on one of the drugs. Indeed, for doxycycline–ciprofloxacin, a region of drug concentrations permitting the growth of doxycycline-sensitive but not resistant strains appears for three very distinct mechanisms of resistance to tetracyclines. It is important to note that the effect observed in the doxycycline–ciprofloxacin combination is unidirectional. Although advantaging the sensitive wild type over a doxycycline-resistant strain, one would not expect it to confer the same benefit over a ciprofloxacin-resistant strain. It would therefore be of considerable interest to employ new multidrug screens[8] to search for reciprocally suppressing drug combinations in which each of the drugs suppresses the effect of the other (Supplementary Fig. S3b). Such drug combinations may block the two single-step mutational paths[28] to complete resistance by imposing selection against resistance to each of the drugs. We emphasize that our work is limited to sublethal drug concentrations, in a controlled environment *in vitro* and that any possible therapeutic implications from these findings are beyond its scope. However, we do hope that these findings may suggest avenues of research into new treatment strategies employing antimicrobial combinations with improved selection against resistance.

METHODS

Media and strains. All experiments were conducted in M63 minimal medium (2 g l^{-1} $(NH_4)_2SO_4$, 13.6 g l^{-1} KH_2PO_4, 0.5 mg l^{-1} $FeSO_4 \bullet 7H_2O$) supplemented with 0.2% glucose, 0.01% casamino acids, 1 mM $MgSO_4$ and 1.5 µM thiamine. Drug solutions were freshly made from powder stocks (doxycycline hyclate, catalogue no. D-9891 (Sigma); ciprofloxacin, catalogue no. 17850 (Fluka); erythromycin, catalogue no. 45673 (Fluka)) and filter-sterilized before each experiment. Strain construction and designations are given in Supplementary Table 3. Assay

strains were grown from single colonies to saturation in supplemented M63 minimal medium. Cell concentrations were measured by plate count, and aliquots were stored in 15% glycerol at $-80\,^{\circ}$C. Fresh aliquots were used for each experiment.

Growth rate assay and MIC line. We used a previously developed luminescence-based assay for measuring exponential growth rates[10,17,18]. Cells containing bacterial luciferase constitutively expressed from the pCS-λ plasmid were grown at $30\,^{\circ}$C (for growth at $37\,^{\circ}$C see Supplementary Fig. S5) in sealed black 96-well microtitre plates (Costar 3792; Corning) in 100 μl of supplemented M63 medium starting from about 100 bacteria per well. For each strain and drug pair, two-dimensional drug concentration gradients were set up in two replicates on 12×16 or 12×30 matrices at concentrations indicated in Supplementary Fig. S6. Emitted light from each well in counts per second (c.p.s.) was recorded by a Topcount NXT plate reader (Perkin-Elmer) at intervals of about 40 min for several days. Growth rates were measured as slopes of log(c.p.s.) over time and thus were insensitive to the absolute intensity per cell. Growth rates were obtained by automatically selecting best least-squares linear fits over sliding intervals within the exponential growth regime (see example in Fig. 2e, and the complete data set in Supplementary Fig. S6). Negative slopes, very low final light levels (c.p.s. < 0.0004 of the assay maximum) and poor fits (root-mean-square > 0.35) were annotated as no data (Supplementary Fig. S6). A response surface was fitted to the average of the replicates by using a smoothing cubic spline, and linearly interpolated isoboles were plotted (with the Matlab functions csaps and contourc). The cubic spline surface reduced noise and did not deviate appreciably from the raw growth data (Supplementary Fig. S7).

The MIC line was defined as the drug concentrations suppressing growth rate to a fixed threshold. To allow comparison with the growth region of the competition assay (comparison between Fig. 2 and Fig. 3), this threshold was defined as the slowest growth rate detectable by optical density at the end of the competition assay. Optical density (absorbance; A) detection in the competition assay requires a roughly 20-fold increase in cell count over 24 h, corresponding to a doubling rate of about $0.17\,h^{-1}$, or 14% of the maximal growth rate in drug-free medium ($1.24\,h^{-1}$). It should still be noted that the values for growth rates and MICs depend on culture conditions, which differ somewhat between the growth (Fig. 2) and competition (Fig. 3) assays.

Competition assay. Our competition assay was derived from that of ref. 23. Sensitive (Wyl and Wcl) and resistant (t17yl and t17cl) strains were labelled with either YFP or CFP on the chromosome, under a strong, constitutive P_{lac} promoter[29,30]. Each strain was labelled with each colour individually for dye-swap control experiments. Cells were introduced in 1:1 ratio, at about 10,000 cells per well, to clear, flat-bottomed, 96-well plates (Costar 3595; Corning) carrying a 12×16 matrix of drug–drug concentrations. Drug concentration ratios were varied along one axis of the matrix, and dilutions ($3^{1/4}$-fold) were performed serially along the other. The plates were covered and then incubated in the dark at $30\,^{\circ}$C for 24 h on a Titramax 1000 shaker (Heidolph) at 600 r.p.m. Numbers of cells expressing each label were determined for wells exhibiting growth ($A_{500} > 0.02$) with a Victor III (Perkin-Elmer) fluorescence plate reader (filters: YFP, HQ500/20x and HQ535/30m; CFP, D436/20x and D480/30m), and additionally by fluorescence-activated cell sorting (FACS) with an LSRII (Becton Dickinson) (CFP, excitation at 405 nm and emission at 450 nm; YFP, excitation at 488 nm and emission at 530 nm; CFP, YFP lower threshold, 100–150 relative light units (RLU)). A surface representing the logarithmic ratio of YFP to CFP cells was plotted over the drug gradient with the use of a linear interpolation between neighbouring data points (Fig. 3 and Supplementary Tables 1 and 2).

Received 18 December 2006; accepted 12 February 2007.

1. Levy, S. B. & Marshall, B. Antibacterial resistance worldwide: causes, challenges and responses. *Nature Med.* **10**, S122–S129 (2004).
2. Golan, D. E. *et al.* (eds). *Principles of Pharmacology: The Pathophysiologic Basis of Drug Therapy* (Lippincott Williams & Wilkins, Philadelphia, 2005).
3. Pillai, S. K., Moellering, R. C. & Eliopoulos, G. M. in *Antibiotics in Laboratory Medicine* (ed. Lorian, V.) 365–440 (Lippincott Williams & Wilkins, Philadelphia, 2005).
4. Yim, G., Wang, H. H. M. & Davies, J. The truth about antibiotics. *Int. J. Med. Microbiol.* **296**, 163–170 (2006).
5. Czaran, T. L., Hoekstra, R. F. & Pagie, L. Chemical warfare between microbes promotes biodiversity. *Proc. Natl Acad. Sci. USA* **99**, 786–790 (2002).
6. Klein, M. & Schorr, S. E. The role of bacterial resistance in antibiotic synergism and antagonism. *J. Bacteriol.* **65**, 454–465 (1953).
7. Lipsitch, M. & Levin, B. R. The population dynamics of antimicrobial chemotherapy. *Antimicrob. Agents Chemother.* **41**, 363–373 (1997).
8. Keith, C. T., Borisy, A. A. & Stockwell, B. R. Multicomponent therapeutics for networked systems. *Nature Rev. Drug Discov.* **4**, 71–78 (2005).
9. Loewe, S. The problem of synergism and antagonism of combined drugs. *Arzneimittelforschung* **3**, 285–290 (1953).
10. Yeh, P., Tschumi, A. I. & Kishony, R. Functional classification of drugs by properties of their pairwise interactions. *Nature Genet.* **38**, 489–494 (2006).
11. Fraser, T. R. The antagonism between the actions of active substances. *BMJ* **2**, 485–487 (1872).
12. Chopra, I. & Roberts, M. Tetracycline antibiotics: Mode of action, applications, molecular biology, and epidemiology of bacterial resistance. *Microbiol. Mol. Biol. Rev.* **65**, 232–260 (2001).
13. Walsh, C. *Antibiotics: Actions, Origins, Resistance*, 335 (American Society for Microbiology Press, Washington DC, 2003).
14. Poole, K. Efflux-mediated antimicrobial resistance. *J. Antimicrob. Chemother.* **56**, 20–51 (2005).
15. Hillen, W. & Berens, C. Mechanisms underlying expression of Tn10 encoded tetracycline resistance. *Annu. Rev. Microbiol.* **48**, 345–369 (1994).
16. Singer, M. *et al.* A collection of strains containing genetically linked alternating antibiotic-resistance elements for genetic-mapping of *Escherichia coli*. *Microbiol. Rev.* **53**, 1–24 (1989).
17. Bjarnason, J., Southward, C. M. & Surette, M. G. Genomic profiling of iron-responsive genes in *Salmonella enterica* serovar typhimurium by high-throughput screening of a random promoter library. *J. Bacteriol.* **185**, 4973–4982 (2003).
18. Kishony, R. & Leibler, S. Environmental stresses can alleviate the average deleterious effect of mutations. *J. Biol.* **2**, 14.1–14.10 (2003).
19. Lenski, R. E., Simpson, S. C. & Nguyen, T. T. Genetic analysis of a plasmid-encoded, host genotype-specific enhancement of bacterial fitness. *J. Bacteriol.* **176**, 3140–3147 (1994).
20. Lenski, R. E. *et al.* Epistatic effects of promoter and repressor functions of the Tn10 tetracycline-resistance operon on the fitness of *Escherichia coli*. *Mol. Ecol.* **3**, 127–135 (1994).
21. Yang, W. R. *et al.* TetX is a flavin-dependent monooxygenase conferring resistance to tetracycline antibiotics. *J. Biol. Chem.* **279**, 52346–52352 (2004).
22. Whittle, G. *et al.* Identification of a new ribosomal protection type of tetracycline resistance gene, *tet*(36), from swine manure pits. *Appl. Environ. Microbiol.* **69**, 4151–4158 (2003).
23. Hegreness, M., Shoresh, N., Hartl, D. & Kishony, R. An equivalence principle for the incorporation of favorable mutations in asexual populations. *Science* **311**, 1615–1617 (2006).
24. Chao, L. Unusual interaction between the target of nalidixic acid and novobiocin. *Nature* **271**, 385–386 (1978).
25. Deeks, S. G. Treatment of anti retroviral-drug-resistant HIV-1 infection. *Lancet* **362**, 2002–2011 (2003).
26. Tsui, W. H. W. *et al.* Dual effects of MLS antibiotics: Transcriptional modulation and interactions on the ribosome. *Chem. Biol.* **11**, 1307–1316 (2004).
27. Drlica, K. & Hooper, D. C. in *Quinolone Antimicrobial Agents* (eds Hooper, D.C. & Rubinstein, E.) 19–40 (ASM Press, Washington DC, 2003).
28. Weinreich, D. M., Delaney, N. F., DePristo, M. A. & Hartl, D. L. Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* **312**, 111–114 (2006).
29. Lutz, R. & Bujard, H. Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I_1-I_2 regulatory elements. *Nucleic Acids Res.* **25**, 1203–1210 (1997).
30. Elowitz, M. B., Levine, A. J., Siggia, E. D. & Swain, P. S. Stochastic gene expression in a single cell. *Science* **297**, 1183–1186 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank U. Alon, N. Q. Balaban, B. Chait, J. Davies, M. Elowitz, Y. Fink, L. Garwin, D. Hartl, M. Hegreness, D. Kahne, M. Kirschner, S. Leibler, R. Lenski, R. Milo, T. Mitchison, R. Moellering, A. Murray, D. Segre', N. Shoresh, S. Walker, C. Walsh, R. Ward and P. Yeh for comments and valuable discussions; and G. Jacoby, N. Shoemaker and G. Wright for their gifts of plasmids. This work was supported partly by a grant from the Human Frontiers Science Program.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to R.K. (roy_kishony@hms.harvard.edu).

LETTERS

Hagfish embryology with reference to the evolution of the neural crest

Kinya G. Ota[1], Shigehiro Kuraku[1] & Shigeru Kuratani[1]

Hagfish, which lack both jaws and vertebrae, have long been the subject of intense interest owing to their position at a crucial point in the evolutionary transition to a truly vertebrate body plan[1–4]. However, unlike the comparatively well characterized vertebrate agnathan lamprey, little is known about hagfish development. The inability to analyse hagfish at early embryonic stages has frustrated attempts to resolve questions with important phylogenetic implications, including fundamental ones relating to the emergence of the neural crest[1,5,6]. Here we report the obtainment of multiple pharyngula-stage embryos of the hagfish species *Eptatretus burgeri* and our preliminary analyses of their early development. We present histological evidence of putative neural crest cells, which appear as delaminated cells that migrate along pathways corresponding to neural crest cells in fish and amphibians[2,7–11]. Molecular cloning studies further revealed the expression of several regulatory genes, including cognates of *Pax6*, *Pax3/7*, *SoxEa* and *Sox9*, suggesting that the hagfish neural crest is specified by molecular mechanisms that are general to vertebrates. We propose that the neural crest emerged as a population of de-epithelialized migratory cells in a common vertebrate ancestor, and suggest that the possibility of classical and molecular embryology in hagfish opens up new approaches to clarifying the evolutionary history of vertebrates.

Tracing the evolutionary origin of vertebrates requires careful comparative studies of this group and its closest phylogenetic relatives. The embryonic development of non-vertebrate chordates has

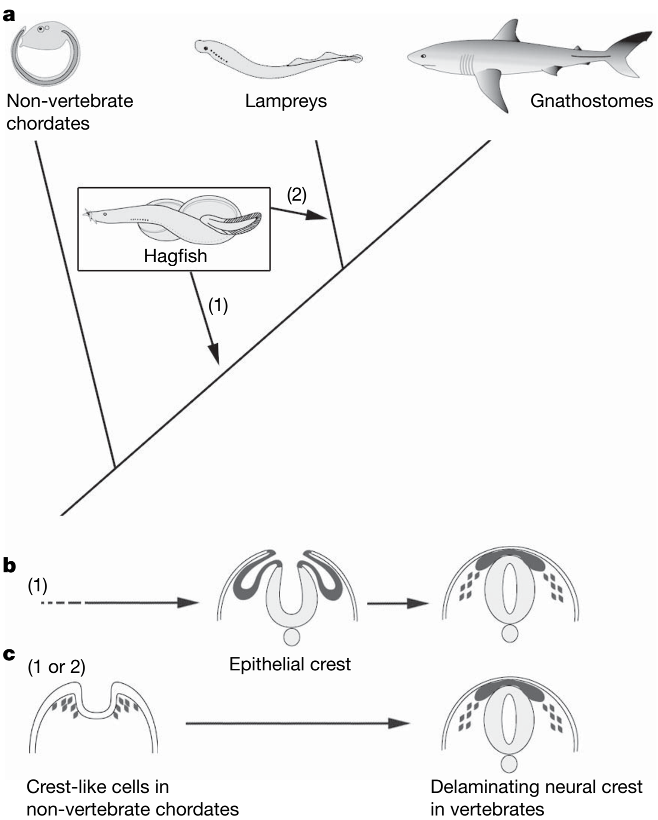


Figure 1 | Evolution of vertebrates and neural crest. **a**, The hagfish may be either the sister group of all the other vertebrates (1), or that of lampreys (2). **b**, **c**, Hypotheses on neural crest evolution. **b**, The former hypothesis (1) would agree with the scenario that neural crest evolution had an intermediate epithelial state. **c**, Alternatively, the crest might have already been established as a population of delaminating cells in the common ancestor to all vertebrates, and this would be coherent with both hypotheses (1 or 2).

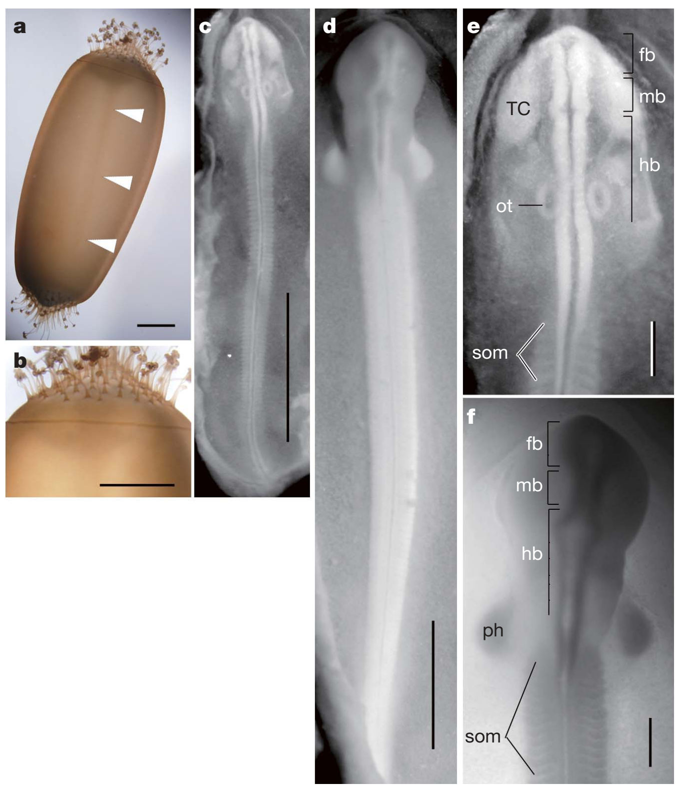


Figure 2 | Embryos of *Eptatretus burgeri*. **a**, **b**, An egg with a 14.3-mm pharyngula-stage embryo inside (arrows). **b**, Higher magnification of the head. **c**, A 7.4-mm embryo, corresponding to the late neurula stage. **d**, A 14.3-mm embryo, corresponding to the early pharyngula stage. **e**, The head of the 7.4-mm embryo. **f**, The head of the 14.3-mm embryo. fb, forebrain; hb, hindbrain; mb, midbrain; ot, otic pit; ph, pharyngeal wall; som, somites; TC, trigeminal crest cells. Scale bars, 5 mm (**a**–**d**); 1 mm (**e**, **f**).

[1]Laboratory for Evolutionary Morphology, Center for Developmental Biology, RIKEN, Kobe 650-0047, Japan.

been extensively studied, as has that of the agnathan (jawless vertebrates) lampreys; however, the development of hagfish, another agnathan group, remains poorly understood, largely owing to the lack of access to embryos[1,2,4,12]. It has been more than one-hundred years since the first series of *Bdellostoma stouti* embryos was obtained[13], and limited attempts have been made since, yielding only a few embryos at later stages in development[4]. The phylogenetic relationship between hagfish, lampreys and gnathostomes (jawed vertebrates) is unresolved, but the hagfish, which lack vertebrae, are often located by morphologists and physiologists as a sister group to the vertebrates (Fig. 1a)[1,2,6]. If this phylogenetic assignment is correct, hagfish embryos might presumably exhibit primitive features not found in lampreys, meaning that the study of this taxon might provide new insights into the evolutionary pathways that led to the emergence of vertebrates.

The ability to make histological observations of hagfish embryos opens up particularly intriguing possibilities for developing a better understanding of the origins of two vertebrate-specific innovations: the ectodermal placode and the neural crest[14]. The presence of placodes has already been suggested in the early hagfish pharyngula[3,15]. As for the crest, it has been suggested that in hagfish this develops as an epithelial pocket arising between the surface ectoderm and neurectoderm, and not as a population of delaminated migrating cells, as is the case in vertebrate embryos (Fig. 1b)[16]. If this model is correct, the delaminated crest would be specific to lampreys and gnathostomes, and absent in hagfish. Even if, as suggested by recent molecular data[17–19], lampreys and hagfish do form a monophyletic group (cyclostomes), analyses of hagfish embryos will still help to define more clearly gnathostome-specific features not shared by cyclostomes. Thus, whichever model is correct, the ability to study hagfish embryos is an absolute prerequisite to developing a better understanding of their true phylogenetic position.

We chose to study a shallow-water hagfish species, *Eptatretus burgeri*, because, in contrast to more deep-water-dwelling species, we felt

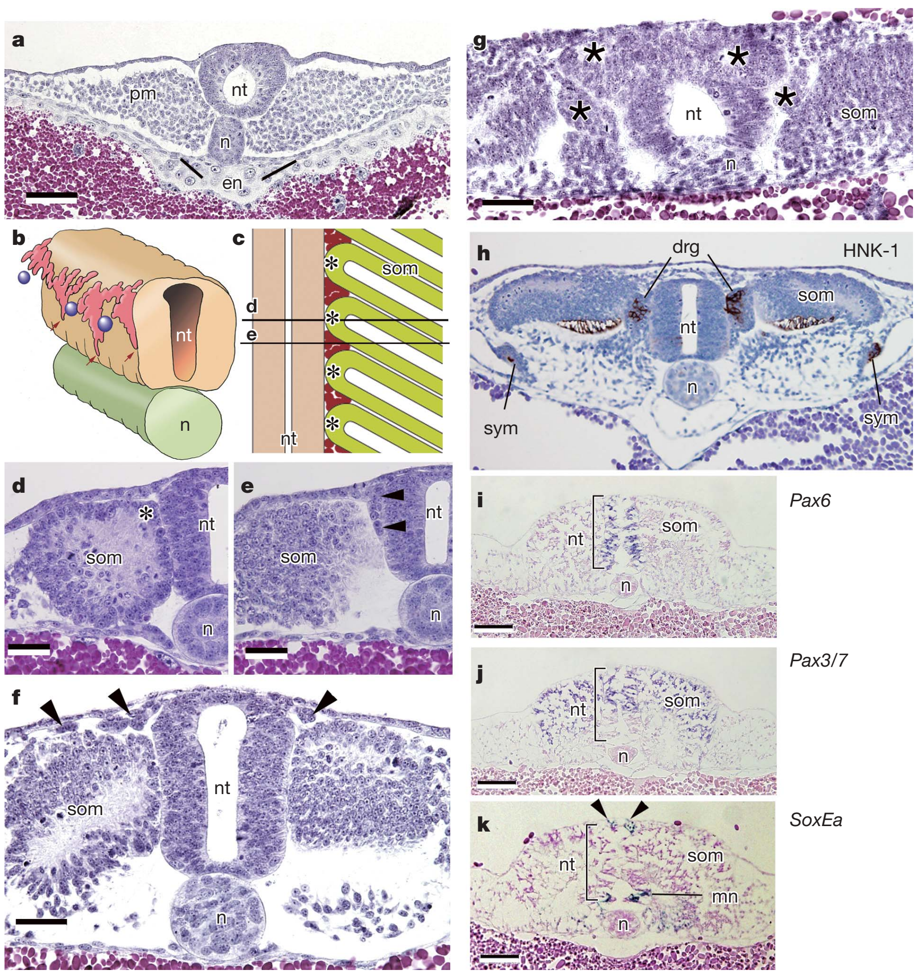


Figure 3 | Neural crest cells in *E. burgeri*. **a**–**e**, Putative neural crest cells in hagfish embryos. **a**–**f**, A 7.4-mm embryo. **a**, No neural crest cells are found in the caudal-most region of the trunk. **b**, Illustration of neural crest cell distribution in the same embryo on the basis of three-dimensional reconstruction from the histological sections; seen from the anterior oblique view. Red arrows indicate the ventrally migrating crest cells occurring at the same intervals as the apices of somites shown by purple circles. **c**, Schematic representation of a horizontal section of the embryo to show the planes of sections **d** and **e**. **d**, A transverse section at the mid-somite level. No putative crest cells are found between the somite (som) and the neural tube (nt). Asterisks in **c** and **d** indicate the apex of each somite. **e**, A transverse section at the intersomitic level. Putative crest cells (arrowheads) are filling in a space between the somite and the neural tube. **f**, Slightly rostral to **a**, showing putative crest cells (arrowheads) beneath the ectoderm. **g**, Transverse section of a 13.4-mm embryo fixed *in toto*. Neuroepithelial pockets (asterisks) appear next to the neural tube. **h**, A 14.3-mm embryo immunostained with HNK-1 antibody. Differentiated neurons including dorsal root ganglia (drg) and sympathetic ganglia (sym), as well as a medial cell population of somites, are stained. **i**–**k**, Expression of the indicated hagfish cognates of the genes involved in vertebrate neural crest specification in the trunk level of a 13.1-mm embryo. *Pax6* is expressed in the entire neural tube (**i**); *Pax3/7* in its dorsal part (**j**). *Pax3/7* is also expressed in the dorsal somites as in gnathostome embryos. **k**, *SoxEa* is expressed in the putative motor neurons (mn) and the pre-migratory neural crest (arrowheads), but not in the migrating crest cells. Brackets indicate the dorsoventral extent of the neural tube. drg, dorsal root ganglia; en, endoderm; mn, motor neurons; n, notochord; nt, neural tube; pm, paraxial mesoderm; som, somite; sym, sympathetic ganglia. Scale bars, 100 μm.

it would be easier for us to recreate their natural habitat in the aquarium[4]. We prepared an aquarium tank in which a large number of males and brood females were kept at low temperature, in an attempt to reproduce the hagfish spawning environment[4]. The animals deposited a total of ninety-two eggs, among which we found seven developing embryos visible through the eggshells in the period from five to seven months after deposition (Fig. 2a, b). These embryos were at four different developmental stages (Supplementary Table 1), confirming the finding that hagfish embryogenesis is asynchronous[13]. It is worth noting that *E. burgeri* takes much longer to hatch than a previous estimation[13] of two months for *B. stouti*, leading us to suspect that some of the eggs previously obtained by other laboratories may have been discarded before development had begun.

To observe the embryos in greater detail, we tested different methods of fixation and compared their histological and whole-mount appearance. As found in the study of ref. 13, which fixed the embryos *in toto* (encapsulated in shells), removal of the eggshell seemed necessary to avoid distortion of specimens. Distortion was enormous in the neural tube of the specimens of ref. 13 (Supplementary Fig. 1), probably owing to unequal swelling and shrinkage of various tissues in a confined space. By fixing an embryo *in toto*, we were able to reproduce the neural tube distortion, and found epithelial pockets lateral to the neural tube, such as were described previously[16] on examination of the specimen of ref. 13 (Fig. 3g; see also Supplementary Fig. 1a).

When we fixed 14.3-mm and 7.4-mm embryos by removing the eggshells, we found well organized brain primordia reminiscent of those in other vertebrate embryos (Fig. 2c–f). On sectioning, we did not find any epithelial pockets, but observed putative migrating neural crest cells at the trunk level (Fig. 3e, f). The crest cells appeared to have delaminated in close proximity to the dorsal neural tube, by a process that seemed to proceed in an anterior to posterior direction (Fig. 3a, e, f), and populated in a segmental pattern associated with somites, as is typical of vertebrate crest cells (Fig. 3b–e)[2,11]. Unlike in amniote embryos[2,11], none of the crest cells migrated into the somites. We performed immunostaining with the monoclonal antibody HNK-1, which is known to recognize migrating crest cells in some vertebrates[11]; however, although this monoclonal antibody did not detect any crest cells, it did label the differentiated dorsal root ganglia that developed in an older embryo (Fig. 3h). These ganglia are also located intersomitically[20], as the crest cells are at previous stages.

To investigate the hagfish neural crest at the molecular level, we isolated two *Pax* genes (*Pax6* and *Pax3/7*) as markers for regionalization of neuroepithelium[21,22], and a *Snail* gene homologue (*SnailA*) and two *SoxE* genes (*SoxEa* and *Sox9*) as candidate neural crest markers[23–25] (Supplementary Figs 2–5). We conducted *in situ* hybridization and found that the expression patterns of *Pax6* and *Pax3/7* were identical to those in the gnathostome neural tube (Fig. 3i, j). Although *SnailA* was not expressed in the putative crest (Supplementary Fig. 6), both the *SoxE* genes (*SoxEa* and *Sox9*) were detected strongly in the neural crest at the trunk level (Figs 3k and 4). In the whole-mount embryo, *Sox9* was also expressed in the segmentally arranged crest cells (Fig. 4), corresponding to the distribution patterns of the putative crest cells in the histological sections (Fig. 3b–e).

These findings indicate that the genetic programmes that specify both the neural crest cells and the overall embryonic architecture of the hagfish resemble those of the typical vertebrate scheme. We have also not detected any feature that would be present in other chordates but not in vertebrates. The absence of *Snail* transcripts in the hagfish neural crest may indicate either an evolutionary change in gene regulation[26] or the presence of another *Slug*-like paralogue in the hagfish; further phylogenetic studies will be needed to resolve this question. Given this new histological and genetic evidence, we suggest that the neural crest probably existed as a population of delaminating and migrating cells in the common ancestor of the entire vertebrate clade, and thus that its origin should be sought in non-vertebrate chordates (Fig. 1a, c). Notably, this model is consistent with the recent finding of 'putative crest cells' in tunicate embryos[27]. These cells not only express a similar set of genes compared to those characteristic of vertebrate neural crest cells, but they also migrate.

Our data indicate that these embryological features, which have been considered to be specific to vertebrates, were in fact already present in the common ancestor of hagfish, lampreys and gnathostomes, and that, given the time of the divergence between hagfish and other vertebrates, their origin could date back to 500 million years ago (Cambrian period)[28,29]. In addition to the origin of neural crest, a number of other questions surrounding hagfish embryogenesis remain, including the origin and development of ectodermal placodes, thyroid and adenohypophysis, the configuration of the cephalic endoderm, and the origin of direct development in their life cycle[4,12,13,30]. Although the present study of hagfish development does not provide conclusive evidence towards resolving the phylogenetic position of this taxon (Fig. 1a), this organism nonetheless remains a key model for exploring vertebrate evolution, and, as we have shown, is now available for use in molecular embryological approaches to address fundamental questions in evolutionary developmental biology.

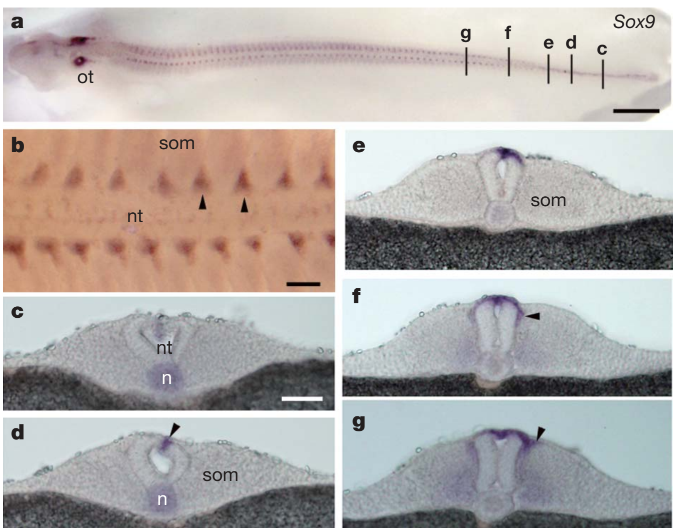


Figure 4 | *Sox9* expression in the neural crest of *E. burgeri*. **a**, **b**, A whole-mount embryo hybridized with a *Sox9* probe. *Sox9* is expressed in otocyst (ot) and neural crest cells populating between somites (som), as in gnathostome embryos. Lines indicate the levels of sections shown in **c**–**g**. **b**, Enlargement of the trunk level. *Sox9*-positive neural crest cells (arrowheads) are predominantly located between intersomitic spaces. **c**–**g**, Transverse sections of the same embryo, shown from caudal to rostral levels. *Sox9* is expressed in the notochord (n) and the dorsal neural tube (nt), representing the pre-migratory neural crest at the caudal-most level (**c** and **d**), and also in the migrating crest cells (arrowheads) at more rostral levels (**e**–**g**). Scale bars, 1 mm (**a**); 100 μm (**b**–**g**).

METHODS

Sample collection and aquarium maintenance. Adult males and females of *E. burgeri* were collected using eel traps from a depth of 25–100 m in the Japan Sea off Shimane prefecture from September to October 2005. After the hagfish were transferred to a laboratory aquarium (at 16 °C) and sexed by manipulation, 25 individuals (13 males and 12 females) were maintained in the aquarium tank lined with some potentially favourable substrates, including fine-grain sands and oyster shells (1,000 l, 16–17 °C), from 10 October 2005 to 30 April 2006. Haematoxylin and eosin staining and HNK-1 immunostaining were performed by standard protocols. Fragments of *Pax6*, *Pax3/7*, *SoxEa*, *Sox9* and *SnailA* were amplified by degenerate polymerase chain reaction with reverse transcription and isolated by the TOPO TA cloning kit dual promoter (Invitrogen). *In situ* hybridization was carried out in a Ventana automated machine (Ventana Medical Systems). Detailed protocols are described in Supplementary Information.

Received 5 October 2006; accepted 22 January 2007.
Published online 18 March 2007.

1. Janvier, P. *Early Vertebrates* (Oxford Univ. Press, Oxford, 1996).
2. Hall, B. K. *The Neural Crest in Development and Evolution* (Springer, New York, 1999).
3. Wicht, H. & Northcutt, R. G. Ontogeny of the head of the Pacific hagfish (*Eptatretus stouti*, Myxinoidea): development of the lateral line system. *Phil. Trans. R. Soc. Lond. B* **349,** 119–134 (1995).
4. Ota, K. G. & Kuratani, S. The history of scientific endeavors towards understanding hagfish embryology. *Zoolog. Sci.* **23,** 403–418 (2006).
5. Holmgren, N. On two embryos of *Myxine glutinosa*. *Acta Zoologica* **27,** 1–90 (1946).
6. Løvtrup, S. *The Phylogeny of Vertebrata* (Wiley, New York, 1977).
7. Bronner-Fraser, M. Environmental influences on neural crest cell migration. *J. Neurobiol.* **24,** 233–247 (1993).
8. Collazo, A., Bronner-Fraser, M. & Fraser, S. E. Vital dye labelling of *Xenopus laevis* trunk neural crest reveals multipotency and novel pathways of migration. *Development* **118,** 363–376 (1993).
9. Eisen, J. S. & Weston, J. A. Development of the neural crest in the zebrafish. *Dev. Biol.* **159,** 50–59 (1993).
10. McCauley, D. W. & Bronner-Fraser, M. Neural crest contributions to the lamprey head. *Development* **130,** 2317–2327 (2003).
11. Le Douarin, N. & Kalcheim, C. *The Neural Crest* (Cambridge Univ. Press, Cambridge, 1999).
12. Gorbman, A. Hagfish development. *Zoolog. Sci.* **14,** 375–390 (1997).
13. Dean, B. On the embryology of *Bdellostoma stouti*. A general account of myxinoid development from the egg and segmentation to hatching. In *Festschrift zum 70ten Geburststag Carl von Kupffer* 220–276 (Gustav Fischer, Jena, 1899).
14. Gans, C. & Northcutt, R. G. Neural crest and the origin of vertebrates: A new head. *Science* **220,** 268–274 (1983).
15. Wicht, H. & Tusch, U. in *The Biology of Hagfish* (eds Jørgensen, J. M., Lomholt, J. R., Weber, R. E. & Malte, H.) 431–451 (Chapman & Hall, London, 1998).
16. Conel, J. L. The origin of the neural crest. *J. Comp. Neurol.* **76,** 191–215 (1942).
17. Mallatt, J. & Sullivan, J. 28S and 18S rDNA sequences support the monophyly of lampreys and hagfishes. *Mol. Biol. Evol.* **15,** 1706–1718 (1998).
18. Takezaki, N., Figueroa, F., Zaleska-Rutczynska, Z. & Klein, J. Molecular phylogeny of early vertebrates: monophyly of the agnathans as revealed by sequences of 35 genes. *Mol. Biol. Evol.* **20,** 287–292 (2003).
19. Furlong, R. F. & Holland, P. W. Bayesian phylogenetic analysis supports monophyly of ambulacraria and of cyclostomes. *Zoolog. Sci.* **19,** 593–599 (2002).
20. Goodrich, E. S. *Studies on the Structure and Development of Vertebrates* (McMillan, London, 1930).
21. Callaerts, P., Halder, G. & Gehring, W. J. *PAX-6* in development and evolution. *Annu. Rev. Neurosci.* **20,** 483–532 (1997).
22. Goulding, M. D., Chalepakis, G., Deutsch, U., Erselius, J. R. & Gruss, P. Pax-3, a novel murine DNA binding protein expressed during early neurogenesis. *EMBO J.* **10,** 1135–1147 (1991).
23. Cheung, M. & Briscoe, J. Neural crest development is regulated by the transcription factor Sox9. *Development* **130,** 5681–5693 (2003).
24. McCauley, D. W. & Bronner-Fraser, M. Importance of SoxE in neural crest development and the evolution of the pharynx. *Nature* **441,** 750–752 (2006).
25. Nieto, M. A., Sargent, M. G., Wilkinson, D. G. & Cooke, J. Control of cell behavior during vertebrate development by Slug, a zinc finger gene. *Science* **264,** 835–839 (1994).
26. Locascio, A., Manzanares, M., Blanco, M. J. & Nieto, M. A. Modularity and reshuffling of *Snail* and *Slug* expression during vertebrate evolution. *Proc. Natl Acad. Sci. USA* **99,** 16841–16846 (2002).
27. Jeffery, W. R., Strickler, A. G. & Yamamoto, Y. Migratory neural crest-like cells form body pigmentation in a urochordate embryo. *Nature* **431,** 696–699 (2004).
28. Shu, D. G. *et al.* Head and backbone of the Early Cambrian vertebrate *Haikouichthys. Nature* **421,** 526–529 (2003).
29. Kumar, S. & Hedges, S. B. A molecular timescale for vertebrate evolution. *Nature* **392,** 917–920 (1998).
30. Gorbman, A. & Tamarin, A. Early development of oral, olfactory and adenohypophyseal structures of agnathans and its evolutionary implications. In *Evolutionary Biology of Primitive Fishes* (eds Foreman, R. E., Gorbman, A., Dodd, J. M. and Olsson, R.) 165–185 (Plenum, New York, 1985).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank O. Kakitani for the sample collection, H. Nagashima and Y. K. Ohya for technical advice, and R. Ladher and D. Sipp for critical reading of this manuscript. This work was supported by Grants-in-Aid from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

Author Contributions K.G.O. performed sample collection, maintenance of the aquarium tank, molecular cloning and *in situ* hybridization. S. Kuraku was particularly engaged in isolation of *Snail* family genes and performed phylogenetic analyses. S. Kuratani operated on the hagfish embryos and conducted histological analysis. K.G.O. and S. Kuratani wrote the manuscript. All of the authors discussed the results and commented on the manuscript.

Author Information Sequences for *Pax6*, *Pax3/7*, *Snail*, *SoxEa* and *Sox9* from *E. burgeri* are deposited in DNA Data Bank of Japan (DDBJ) under accession numbers AB270704, AB270703, AB288229, AB288230 and AB270702, respectively. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to S. Kuratani (saizo@cdb.riken.jp).

Vol 446 | 5 April 2007 | doi:10.1038/nature05674

LETTERS

Myc deletion rescues Apc deficiency in the small intestine

Owen J. Sansom[1], Valerie S. Meniel[2], Vanesa Muncan[3], Toby J. Phesse[2], Julie A. Wilkins[1], Karen R. Reed[2], J. Keith Vass[1], Dimitris Athineos[1], Hans Clevers[3] & Alan R. Clarke[2]

The *APC* gene encodes the adenomatous polyposis coli tumour suppressor protein, germline mutation of which characterizes familial adenomatous polyposis (FAP), an autosomal intestinal cancer syndrome[1]. Inactivation of *APC* is also recognized as the key early event in the development of sporadic colorectal cancers[2,3], and its loss results in constitutive activity of the β-catenin–Tcf4 transcription complex[3]. The proto-oncogene *c-MYC* has been identified as a target of the Wnt pathway in colorectal cancer cells *in vitro*[4], in normal crypts *in vivo*[5] and in intestinal epithelial cells acutely transformed on *in vivo* deletion of the *APC* gene[6]; however, the significance of this is unclear. Therefore, to elucidate the role Myc has in the intestine after Apc loss, we have simultaneously deleted both *Apc* and *Myc* in the adult murine small intestine. Here we show that loss of Myc rescued the phenotypes of perturbed differentiation, migration, proliferation and apoptosis, which occur on deletion of *Apc*. Remarkably, this rescue occurred in the presence of high levels of nuclear β-catenin. Array analysis revealed that Myc is required for the majority of Wnt target gene activation following Apc loss. These data establish Myc as the critical mediator of the early stages of neoplasia following Apc loss.

We have previously used a conditional gene deletion approach to study the immediate effects of loss of Apc on differentiation, proliferation and migration in the murine small intestine[6]. These studies have shown that loss of Apc leads to unrestricted proliferation within the intestinal crypt, resulting in multiple changes, including an increase in BrdU incorporation, MCM staining and dramatically enlarged crypts at four days post Apc loss[6]. Similar analyses of Myc deficiency have shown remarkably little effect of gene deletion on intestinal proliferation and apoptosis over a five day time course[7,8]. To examine how Myc affected the phenotypes associated with Apc loss, we intercrossed $AhCre^{+}Apc^{fl/fl}$ mice with $Myc^{fl/fl}$ mice. *AhCre* mice carry a *Cre* transgene that is under control of the cytochrome p450 1A1 (*CYP1A1*) promoter. Mice were given three injections of β-napthoflavone on a single day, which induces Cre recombinase activity within the intestine. The mice were analysed four days after the first injection. Staining for β-galactosidase of intestinal whole-mounts from $AhCre^{+}Apc^{fl/fl}Myc^{fl/fl}$ mice carrying the *ROSA26* floxed stop cassette *lacZ* reporter showed near 100% recombination of the reporter allele, indicating retention of double-mutant *Apc*-, *Myc*-deficient crypts (Supplementary Fig. 1a). This highlighted that Cre-mediated deletion of both *Myc* and *Apc* was not immediately deleterious to intestinal enterocytes. To confirm that this protocol produced crypts that were both *Apc*- and *Myc*-deficient we performed immunohistochemistry for Myc and nuclear β-catenin (a downstream marker of Apc loss). Figure 1a–c shows loss of Myc positivity in crypts by immunostaining and Fig. 1d–f shows nuclear β-catenin localization consistent with loss of both Apc and Myc in $AhCre^{+}Apc^{fl/fl}Myc^{fl/fl}$ intestines.

To confirm co-incident Apc and Myc loss, we performed *in situ* hybridization for *Apc* and immunohistochemistry for Myc and β-catenin in serial sections and saw loss of both *Apc* and Myc and the appearance of nuclear β-catenin (Supplementary Fig. 2). Indeed, scoring serial sections revealed combined gene loss in 88.5% (±5% s.d.) of crypts ($n = 3$ mice, 100 crypts scored per mouse). Finally, to show loss of Apc protein, we performed immunofluorescence for Apc (Supplementary Fig. 3). Histological analysis showed that wild-type and doubly deficient crypts were morphologically indistinguishable (Fig. 2a–c). Gross changes associated with Apc deficiency such as the increased crypt cellularity (Fig. 2g), perturbed differentiation (Fig. 2d–f) and an increased number of mitotic figures (Fig. 2h) were absent in the double mutants. To investigate further proliferative changes, we examined levels of BrdU incorporation (Fig. 3a–c) and Ki67 staining (data not shown). In both cases the double-mutant mice did not differ significantly from wild-type mice (Mann–Whitney U test; $P = >0.4$), and showed lower indices than *Apc* null tissue and a restoration of the normal proliferation zone (Mann–Whitney U test; $P = 0.04$) (Fig. 3d–f). We next scored the capacity of double-mutant cells to migrate along the crypt villus axis by comparing the position of BrdU positive cells at 2 and 24 h after exposure. Remarkably, unlike Apc-deficient cells, migration in the double-mutant mice was normal (Fig. 3f). This rescue explains the presence of cells with nuclear β-catenin in the double-mutant villi (see Fig. 1f).

Levels of nuclear β-catenin remained high in the doubly deficient intestines, providing an ideal opportunity to dissect out Myc-dependent and -independent targets of the Wnt pathway. We therefore performed microarray analysis with the affymetrix MOE430_2 chip using RNA from $AhCre^{+}Apc^{fl/fl}Myc^{fl/fl}$, $AhCre^{+}Apc^{fl/fl}Myc^{+/+}$ and $AhCre^{+}Apc^{+/+}Myc^{+/+}$ intestines at four days after gene deletion[6]. The data were analysed using fold change, rank product[9] and SAM[10] (statistical analysis of microarrays) analysis with a false discovery rate (FDR) of no more that 5%. These analyses indicated significant changes between the three genotypes, listed in Supplementary Tables 1–3. Following only Apc loss, SAM analysis using a 5% FDR identified 58 genes that had previously been defined by the literature as Wnt target genes[11] (many are shown at the following website: http://www.stanford.edu/~rnusse/pathways/targets.html). Similar, overlapping gene sets were identified by fold change or rank product analysis (46 known Wnt target genes were increased by 2-fold or greater; 51 were identified according to rank product analysis). We next examined the expression of these genes in the $Apc^{fl/fl}$ $Myc^{fl/fl}$ double mutants and found that additional deficiency of Myc blocked the upregulation of the majority of these Wnt target genes (see Supplementary Table 4). Sixty-two per cent of genes (36/58) identified by SAM analysis were significantly downregulated in the double

[1]The Beatson Institute, Garscube Estate, Glasgow G61 1BD, UK. [2]Cardiff School of Biosciences, Cardiff University, Cardiff CF10 3US, UK. [3]Hubrecht Laboratory, Nederlands Institute for Developmental Biology, Upsalalaan 8, 3584 CT Utrecht, The Netherlands.

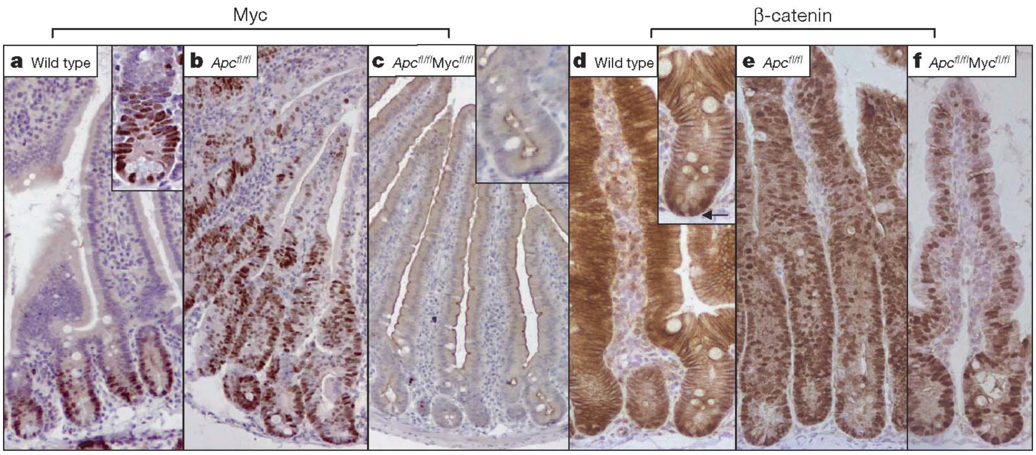


Figure 1 | Efficient deletion of both *Apc* and *Myc* from intestinal enterocytes. **a**–**c**, Myc shown by immunohistochemistry. **a**, Myc in wild-type crypts (*AhCre*$^+$*Apc*$^{++}$*Myc*$^{++}$ mice) at four days post Cre induction. Inset shows nuclear Myc expression at the base of the crypt. **b**, Upregulation of Myc following Apc loss in Apc-deficient crypts (*AhCre*$^+$*Apc*$^{fl/fl}$*Myc*$^{+/+}$ mice) at four days post Cre induction. The immunohistochemistry shows nuclear Myc in every Apc-deficient cell in the crypt. **c**, Loss of Myc protein in double-mutant crypts (*AhCre*$^+$*Apc*$^{fl/fl}$*Myc*$^{fl/fl}$) at four days following Cre induction. Inset shows that no cells express Myc within the crypt. **d**–**f**, Upregulation of β-catenin in double-mutant (*AhCre*$^+$*Apc*$^{fl/fl}$*Myc*$^{fl/fl}$) intestines shown by immunohistochemistry. **d**, β-catenin is localized at the cell edge of nearly all intestinal enterocytes of wild-type crypts (*AhCre*$^+$*Apc*$^{+/+}$*Myc*$^{++}$ mice) at four days post Cre induction. Arrow points to the occasional cell which has nuclear β-catenin at the base of a normal crypt. **e**, Nuclear localization of β-catenin in every cell within Apc-deficient crypts (*AhCre*$^+$*Apc*$^{fl/fl}$*Myc*$^{+/+}$ mice) at four days post Cre induction. **f**, Nuclear localization of β-catenin in every cell within double-mutant crypts (*AhCre*$^+$*Apc*$^{fl/fl}$*Myc*$^{fl/fl}$ mice) at four days post Cre induction. Because β-catenin adopts only a nuclear localization throughout the crypt if Apc is deleted, these data argue that in the double-mutant crypts both *Apc* and *Myc* are deleted. Magnifications, ×400.

knockouts; 27/46 (59%) were downregulated as assessed by fold change; and 32/51 (62%) were significantly downregulated according to rank product analysis. These genes include *axin2*, *tcf1*, *tiam1* (ref. 12) and *Sox17*, and thus are identified as Myc-dependent (Supplementary Table 4). In contrast, a smaller percentage of these genes (including *CD44*, *Cyclin D2* and *Troy*) were unaffected by the additional loss of Myc, reflecting Myc independence. We confirmed a subset of these changes by real time quantitative reverse transcriptase (qRT)–PCR (Supplementary Fig. 4).

Thus, we define, in an *in vivo* setting, a subset of Wnt target genes for which Myc is essential. This subset of target genes is critical in imposing the Apc-deficient phenotype. For example, we have previously shown that perturbation of the EphB/EphrinB system, following Apc loss, results in mislocalization of Paneth cells[6,13]. In the double mutants, these Tcf4 target genes were no longer transcriptionally elevated, and consistent with this, we observed normal localization of the Paneth cells in the double mutants (Fig. 3g–j). Critically, the Wnt target genes that we identify as Myc-independent

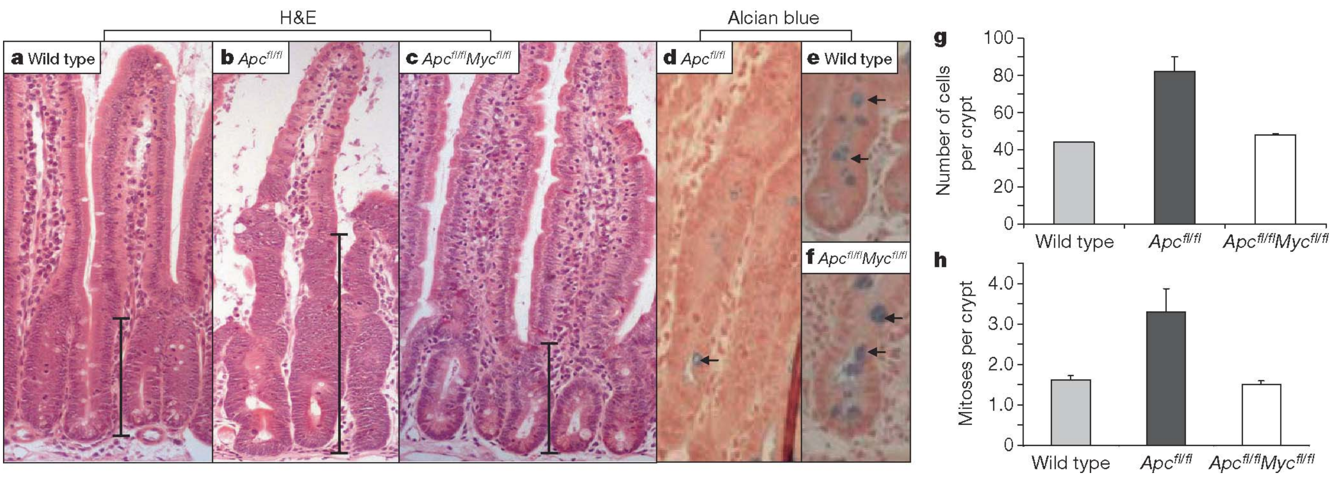


Figure 2 | Additional loss of Myc causes Apc-deficient crypts to resemble wild-type crypts. **a**–**c**, Haematoxylin and eosin (H&E)-stained sections for wild-type (*AhCre*$^+$*Apc*$^{+/+}$*Myc*$^{+/+}$) (**a**), Apc-deficient (*AhCre*$^+$*Apc*$^{fl/fl}$*Myc*$^{+/+}$) (**b**) and double-mutant Apc- and Myc-deficient (*AhCre*$^+$*Apc*$^{fl/fl}$*Myc*$^{fl/fl}$) (**c**) intestines. Note the enlarged crypts in the Apc-deficient mice (**b**) are no longer present in double-mutant intestines (**c**). **d**–**f**, Restored goblet cell number in double-mutant intestines. Alcian blue-stained section for Apc-deficient (*AhCre*$^+$*Apc*$^{fl/fl}$*Myc*$^{+/+}$) (**d**), wild-type (*AhCre*$^+$*Apc*$^{+/+}$*Myc*$^{+/+}$) (**e**), and double-mutant (*AhCre*$^+$*Apc*$^{fl/fl}$*Myc*$^{fl/fl}$) (**f**) intestines. Note the reduction in goblet cells that occurs in Apc-deficient crypts does not occur in double-mutant crypts (Mann–Whitney test, $P = 0.04$, $n = 3$). Arrows point to 'blue' goblet cells. **g**, Crypt cellularity is not increased in double- mutant crypts. Graph showing average cell number per crypt. Error bars, s.d. Unlike Apc-deficient crypts, which have a significantly increased size compared with wild type crypts ($P = 0.04$, Mann–Whitney test, $n = 3$), no difference in size was observed between wild-type and double-mutant crypts (Mann–Whitney test, $P = 0.4$). **h**, Numbers of mitotic figures per crypt are not increased in double-mutant crypts. Graph showing average number of mitoses per crypt. Error bars, s.d. Unlike Apc-deficient crypts, which have a significantly increased number of mitotic figures per crypt compared with wild-type crypts ($P = 0.04$, Mann–Whitney test, $n = 3$), no significant difference was observed between wild-type and double-mutant crypts (Mann–Whitney test, $P = 0.303$). In all figures n refers to the number of mice analysed. Error bars, s.d. In **a**–**c**, scale bars show size of crypt; magnification of **d**–**f**, ×600.

are insufficient to impose the phenotypes associated with Apc deficiency.

In the double mutants, a number of genes associated with DNA and RNA replication were no longer transcriptionally upregulated (for example, DNA polymerase ε, *MCM2*, *MCM7*, *Brf1*, RNA polymerase I subunits, RNA polymerase II, *eif4γ*, and the rRNA genes (*RpS11–13*, *S18* and *L13*) . Our data are therefore consistent with numerous reports of Myc driving transcription of genes associated with proliferation[14]; however, we cannot rule out the possibility that these changes simply reflect the reduced proliferation that we observe.

Previous studies have postulated that Myc may facilitate proliferation through repression of p21 (ref. 5). This is unlikely to be a direct effect within normal intestinal crypts because there is no induction of p21 and no obvious G1 arrest when Myc is deleted[7,8]. However, repression of p21 by Myc was demonstrated in colorectal cancer cell lines mutant for Apc, raising the possibility that repression of p21 may only be important in the context of activated Wnt signalling. To address this, we performed immunohistochemistry for p21 (Fig. 3k–m) and found p21 to be upregulated within the crypts of double-mutant $Apc^{fl/fl}Myc^{fl/fl}$ mice. This argues that the loss of the crypt progenitor-cell-like phenotype in double-mutant mice may be partially due to the inability to repress p21.

Previous reports have suggested that Myc can integrate and balance different survival signals. In certain contexts, Myc expression can lead to proliferation, whereas in other contexts, activation of Myc drives apoptosis[15,16]. Given the failure to see increased proliferation in the double mutants, we next investigated levels of apoptosis in the $AhCre^{+}Apc^{fl/fl}Myc^{fl/fl}$ intestines. Previously, we have shown that loss of Apc increases the apoptotic index in $AhCre^{+}Apc^{fl/fl}$ mice. In the double-mutant $AhCre^{+}Apc^{fl/fl}Myc^{fl/fl}$ intestines this increase was completely blocked (wild type, 0.16% ± 0.05; $Apc^{fl/fl}$, 7.3% ± 0.6; $Apc^{fl/fl}Myc^{fl/fl}$, 0.33 ± 0.13; Supplementary Fig. 5a), showing Myc dependency for both apoptosis and proliferation, following activation of the Wnt pathway. To investigate further the link between proliferation and death, we pulse-labelled with BrdU and followed the fate of labelled cells. Within wild-type crypts, there was a 43% increase in the number of labelled cells (as a consequence of division) between 2 and 24 h (Supplementary Fig. 5b). No such increase was observed in the $AhCre^{+}Apc^{fl/fl}Myc^{+/+}$ mice, implying that BrdU-positive cells were being deleted (Fig. 4b). This interpretation is further supported by the observation that the mitotic index was not elevated in the $Cre^{+}Apc^{fl/fl}$ mice (control, 5.83 ± 0.27 s.e.; $Cre^{+}Apc^{fl/fl}$, 6.43 ± 1.77; $P = 0.66$, Mann–Whitney test, $n = 3$), despite the large increase in BrdU labelling, implying that death is

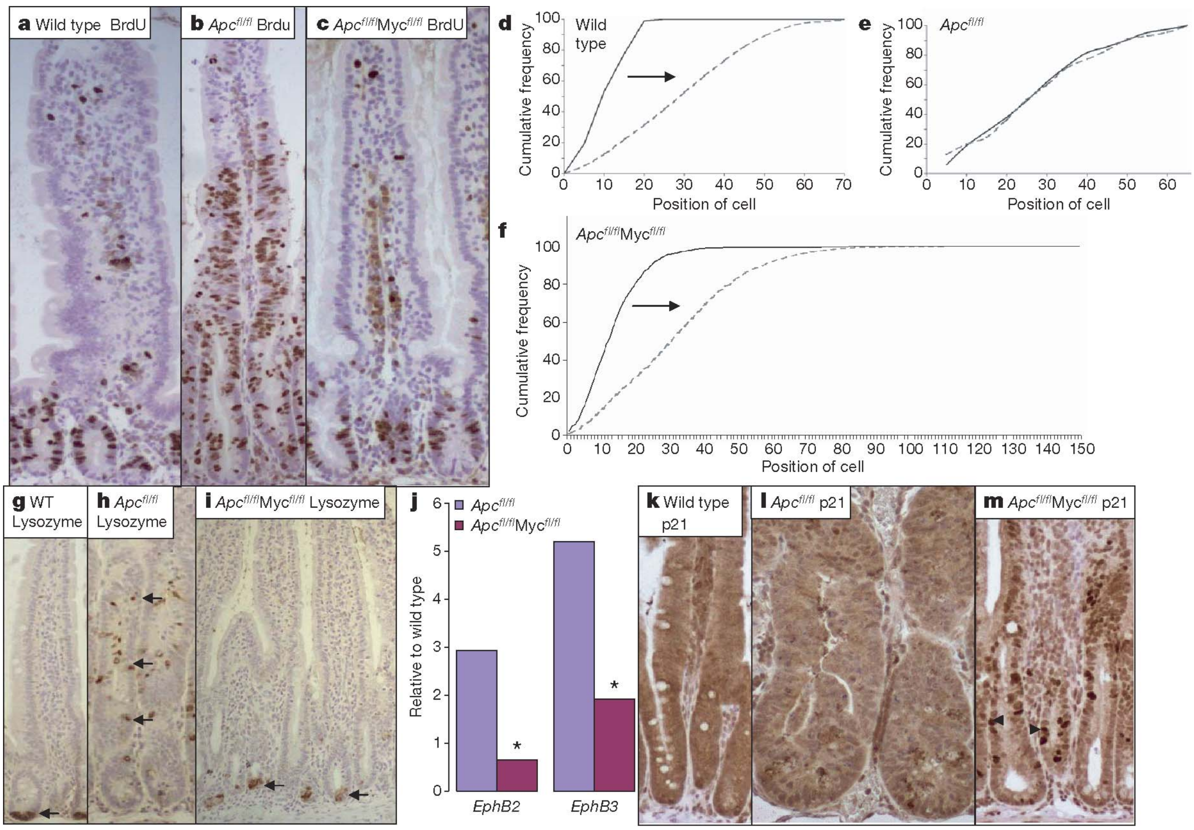


Figure 3 | Additional loss of Myc removes the crypt progenitor-cell-like phenotype of Apc-deficient cells. a–c, Immunohistochemistry for BrdU performed on wild-type ($AhCre^{+}Apc^{+/+}Myc^{+/+}$) (**a**), Apc-deficient ($AhCre^{+}Apc^{fl/fl}Myc^{+/+}$) (**b**) and double-mutant ($AhCre^{+}Apc^{fl/fl}\ Myc^{fl/fl}$) (**c**) intestines, at four days after *Cre* induction and two hours following BrdU injection. Note that wild-type and double-mutant crypts only incorporate BrdU within the crypt, displaying a defined proliferation zone. Apc-deficient crypts have a significantly increased number and percentage of BrdU-labelled cells (Mann–Whitney test, $P = 0.04$). **d–f**, Graphs showing the position of the BrdU-labelled cells 2 h (black solid line) and 24 h (grey dashed line) after labelling. In wild-type ($AhCre^{+}Apc^{+/+}Myc^{+/+}$) crypts, cells migrate up the crypt–villus axis from 2 to 24 h; shown here by cells labelling at higher positions on the crypt–villus axis (arrows; Kolmogorov–Smirnov test, 2 h versus 24 h, $P = 0.01$). Within Apc-deficient ($AhCre^{+}Apc^{fl/fl}Myc^{+/+}$) crypts, cells label throughout the crypt–villus axis at 2 h and there is no movement 24 h later (Kolmogorov–Smirnov test, 2 h versus 24 h, $P = 0.8$). Within double-mutant ($AhCre^{+}Apc^{fl/fl}Myc^{fl/fl}$) crypts, cells migrated up the crypt–villus axis (Kolmogorov–Smirnov test, 2 h versus 24 h, $P = 0.01$). **g–i**, Immunohistochemistry for lysozyme, performed on wild-type (WT; $AhCre^{+}Apc^{+/+}Myc^{+/+}$) (**g**), Apc-deficient ($AhCre^{+}Apc^{fl/fl}Myc^{+/+}$) (**h**) and double-mutant ($AhCre^{+}Apc^{fl/fl}Myc^{fl/f}$) (**i**) intestines. Positions of lysozyme-positive paneth cells are denoted by arrows. **j**, Real time qRT–PCR for the EphB2 and EphB3 receptor mRNAs. Note Paneth cells are no longer mislocalized in double-mutant crypts and expression of the EphB2 and EphB3 receptor mRNAs is significantly lower than in Apc-deficient crypts (*$P = 0.04$, Mann–Whitney test). **k–m**, Immunohistochemistry for p21, performed on wild-type (**k**), Apc-deficient (**l**) and double-mutant ($AhCre^{+}Apc^{fl/fl}Myc^{fl/fl}$)(**m**) crypts. Arrowheads indicate a number of p21-positive cells in double-mutant crypts (**m**). Magnifications, ×400.

occurring in either the G2 or M phases[17]. In accordance with this, many of the apoptotic figures observed were large—a phenomenon previously interpreted to reflect death of 4*n* cells at a G2/M checkpoint[17]. This is also consistent with reports that Myc upregulation drives cells precociously into S phase[18] and β-catenin induces death at the G2/M checkpoint in epidermal keratinocytes *in vitro*[19].

To investigate the long-term fate of doubly deficient cells, we intercrossed $AhCre^{+}Apc^{fl/fl}Myc^{fl/fl}$ mice with mice carrying the *ROSA26 lacZR* allele[20] to report Cre-mediated recombination, and exposed them to dietary β-napthoflavone (0.8 mg ml^{-1}). This approach delivers lower levels of recombination, and allows $AhCre^{+}Apc^{fl/fl}$ mice to survive sufficiently long to develop adenomas[21]. Notably, this strategy also allows an analysis of the interaction between mutant and wild-type cells because approximately 10% of crypts undergo recombination. $AhCre^{+}Apc^{fl/fl}Myc^{+/+}$ and $AhCre^{+}Apc^{fl/fl}Myc^{fl/fl}$ mice were induced and analysed 20 days later. The $AhCre^{+}Apc^{fl/fl}Myc^{+/+}$ mice develop severe intestinal pathology, with thousands of small lesions and multiple adenomas. In contrast, no double-mutant cells remained in $AhCre^{+}Apc^{fl/fl}Myc^{fl/fl}$ mice and there was no pathological change. This was confirmed by immunostaining, which failed to identify any crypts that were Myc deficient or that retained nuclear β-catenin (Supplementary Fig. 6).

We have previously shown that single deficiency of Myc leads to reduced biosynthetic activity and strong selection against Myc-deficient stem cells[8]. We hypothesize that, in the double mutants, surrounding wild-type stem cells outcompete the double-null cells and repopulate the crypts, much as we have seen for other alleles deleted using this strategy[22,23].

In conclusion, we show that the dramatic changes conferred by Apc deficiency are entirely dependent on functional Myc. Furthermore, given that Myc is required for the persistence of Apc-deficient cells, these data show that Myc is absolutely required for the cellular and molecular changes that occur following Apc loss in the murine small intestine.

METHODS

Mouse colonies. All experiments were performed under the UK Home Office guidelines. Outbred male mice from 6–12 weeks of age were used, and they were segregating for the C57BLJ and S129 genomes. The alleles used were as follows: $Apc^{580S\ flox}$ (ref. 6), $Myc^{fl/fl}$ (ref. 8) and *AhCre* (ref. 22).

Tissue isolation and analysis. To induce recombination, mice were given daily intraperitoneal injections of β-napthoflavone (80 mg kg^{-1}) for 3 days and harvested 4 days after the first injection. At this time point, mice were killed and the small intestine removed and flushed with water. Intestines were dissected and analyses for apoptosis, mitosis and S-phase labelling were performed on the proximal intestine, as described in Supplementary Methods. *In situ* hybridization, immunohistochemistry and immunofluorescence were performed as previously described[9,11]. Detailed protocols are given in the Supplementary Methods. To determine the pattern of recombination at the *ROSA26R* reporter locus, intestinal whole-mounts were prepared, fixed and exposed to X-gal substrate using a method previously reported[8,23]. For microarray analysis, RNA was extracted from 3 cm of the small intestine that was located 7 cm from the stomach. For details of RNA extraction, microarray analysis and real time qRT–PCR analysis of targets see Supplementary Methods.

Received 21 September 2006; accepted 9 February 2007.
Published online 21 March 2007.

1. Kinzler, K. W. *et al.* Identification of FAP locus genes from chromosome 5q21. *Science* **253,** 661–665 (1991).
2. Kinzler, K. W. & Vogelstein, B. Lessons from hereditary colorectal cancer. *Cell* **87,** 159–170 (1996).
3. Korinek, V. *et al.* Constitutive transcriptional activation by a β-catenin–Tcf complex in $APC^{-/-}$ colon carcinoma. *Science* **275,** 1784–1787 (1997).
4. He, T. C. *et al.* Identification of c-MYC as a target of the APC pathway. *Science* **281,** 1509–1512 (1998).
5. van de Wetering, M. *et al.* The β-catenin/TCF-4 complex imposes a crypt progenitor phenotype on colorectal cancer cells. *Cell* **111,** 241–250 (2002).
6. Sansom, O. J. *et al.* Loss of Apc *in vivo* immediately perturbs Wnt signaling, differentiation, and migration. *Genes Dev.* **18,** 1385–1390 (2004).
7. Bettess, M. D. *et al.* c-Myc is required for the formation of intestinal crypts but dispensable for homeostasis of the adult intestinal epithelium. *Mol. Cell. Biol.* **25,** 7868–7878 (2005).
8. Muncan, V. *et al.* Rapid loss of intestinal crypts on conditional deletion of the Wnt/Tcf-4 target gene c-Myc. *Mol. Cell. Biol.* **26,** 8418–8426 (2006).
9. Breitling, R., Armengaud, P., Amtmann, A. & Herzyk, P. Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments. *FEBS Lett.* **573,** 83–92 (2004).
10. Tusher, V. G., Tibshirani, R. & Chu, G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl Acad. Sci. USA* **98,** 5116–5121 (2001).
11. Giles, R. H., van Es, J. H. & Clevers, H. Caught up in a Wnt storm: Wnt signaling in cancer. *Biochim. Biophys. Acta* **1653,** 1–24 (2003).
12. Malliri, A. *et al.* The Rac activator Tiam1 is a Wnt-responsive gene that modifies intestinal tumor development. *J. Biol. Chem.* **281,** 543–548 (2006).
13. Batlle, E. *et al.* β-catenin and TCF4 mediate cell positioning in the intestinal epithelium by controlling the expression of EphB/EphrinB. *Cell* **111,** 251–263 (2002).
14. Oskarsson, T. & Trumpp, A. The Myc trilogy: lord of RNA polymerases. *Nature Cell Biol.* **7,** 215–217 (2005).
15. Pelengaris, S., Khan, M. & Evan, G. I. Suppression of Myc-induced apoptosis in β cells exposes multiple oncogenic properties of Myc and triggers carcinogenic progression. *Cell* **109,** 321–334 (2002).
16. Green, D. R. & Evan, G. I. A matter of life and death. *Cancer Cell* **2002,** 19–30 (2002).
17. Merritt, A. J., Allen, T. D., Potten, C. S. & Hickman, J. A. Apoptosis in small intestinal epithelia from *p53*-null mice: evidence for a delayed, p53-independent G2/M-associated cell death after γ-irradiation. *Oncogene* **14,** 2759–2766 (1997).
18. Li, Q. & Dang, C. V. c-Myc overexpression uncouples DNA replication from mitosis. *Mol. Cell. Biol.* **19,** 5339–5351 (1999).
19. Olmeda, D., Castel, S., Vilaro, S. & Cano, A. β-catenin regulation during the cell cycle: implications in G2/M and apoptosis. *Mol. Biol. Cell.* **14,** 2844–2860 (2003).
20. Soriano, P. Generalized *lacZ* expression with the ROSA26 Cre reporter strain. *Nature Genet.* **21,** 70–71 (1999).
21. Sansom, O. J. *et al.* Cyclin D1 is not an immediate target of β-catenin following Apc loss in the intestine. *J. Biol. Chem.* **280,** 28463–28467 (2005).
22. Ireland, H. *et al.* Inducible Cre-mediated control of gene expression in the murine gastrointestinal tract: effect of loss of β-catenin. *Gastroenterology* **126,** 1236–1246 (2004).
23. Hay, T., Patrick, T., Winton, D., Sansom, O. J. & Clarke, A. R. *Brca2* deficiency in the murine small intestine sensitizes to p53-dependent apoptosis and leads to the spontaneous deletion of stem cells. *Oncogene* **24,** 3842–3846 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements. This work was supported by the CR-UK. Thanks to D. Scarborough and Beatson technology services (T. Gilby, M. O'Prey and A. Dawson) for help with histology, M. Bishop for genotyping and K. Vousden for comments. Thanks also to Y. Hey and the PICR for microarray analysis.

Author Contribution: O.J.S., V.S.M., V.M., K.R.R., T.J.P. J.A.W., J.K.V. and D.A. conducted research for the paper. O.J.S., H.C. and A.R.C. wrote the paper. All authors discussed the results and read the paper.

Author Information: Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to O.J.S. (o.sansom@beatson.gla.ac.uk).

Vol 446|5 April 2007|doi:10.1038/nature05652

LETTERS

Semaphorin 7A initiates T-cell-mediated inflammatory responses through α1β1 integrin

Kazuhiro Suzuki[1], Tatsusada Okuno[3], Midori Yamamoto[1], R. Jeroen Pasterkamp[4], Noriko Takegahara[2], Hyota Takamatsu[1], Tomoe Kitao[5], Junichi Takagi[5], Paul D. Rennert[6], Alex L. Kolodkin[7], Atsushi Kumanogoh[1,2] & Hitoshi Kikutani[1]

Semaphorins are axon guidance factors that assist growing axons in finding appropriate targets and forming synapses[1]. Emerging evidence suggests that semaphorins are involved not only in embryonic development but also in immune responses[2]. Semaphorin 7A (Sema7A; also known as CD108)[3,4], which is a glycosylphosphatidylinositol-anchored semaphorin, promotes axon outgrowth through β1-integrin receptors and contributes to the formation of the lateral olfactory tract[5]. Although Sema7A has been shown to stimulate human monocytes[6], its function as a negative regulator of T-cell responses has also been reported[7]. Thus, the precise function of Sema7A in the immune system remains unclear. Here we show that Sema7A, which is expressed on activated T cells, stimulates cytokine production in monocytes and macrophages through α1β1 integrin (also known as very late antigen-1) as a component of the immunological synapse, and is critical for the effector phase of the inflammatory immune response. Sema7A-deficient (*Sema7a*$^{-/-}$) mice are defective in cell-mediated immune responses such as contact hypersensitivity and experimental autoimmune encephalomyelitis. Although antigen-specific and cytokine-producing effector T cells can develop and migrate into antigen-challenged sites in *Sema7a*$^{-/-}$ mice, *Sema7a*$^{-/-}$ T cells fail to induce contact hypersensitivity even when directly injected into the antigen-challenged sites. Thus, the interaction between Sema7A and α1β1 integrin is crucial at the site of inflammation. These findings not only identify a function of Sema7A as an effector molecule in T-cell-mediated inflammation, but also reveal a mechanism of integrin-mediated immune regulation.

We first examined whether Sema7A could bind and stimulate monocytes and macrophages through integrins. THP-1 cells, a human monocytic cell line, were treated with a panel of function-blocking monoclonal antibodies against integrins, and subjected to adhesion assays on plates that were coated with Sema7A fused with the Fc portion of human IgG1 (Sema7A–Fc). Pre-treatment of THP-1 cells with anti-β1 monoclonal antibody strongly inhibited binding of the cells to Sema7A–Fc (Fig. 1a), which is consistent with a previous observation that Sema7A-induced axon growth is blocked by anti-β1 monoclonal antibody[5]. Notably, treatment with anti-α1 monoclonal antibody also resulted in significant inhibition of binding, and blocking both α1 and β1 integrins reduced the number of bound cells to background levels. In contrast, monoclonal antibodies against α2, α3, α4, α5, α6 and αv subunits—all of which form functional heterodimers with the β1 subunit[8]—and monoclonal antibodies recognizing the αvβ3 and αvβ5 complexes had no inhibitory effects. Although it has been proposed that plexin C1 is a Sema7A receptor[9], the effect of Sema7A on neuronal cells is independent of plexin C1 (ref. 5). Anti-plexin C1 monoclonal antibody did not affect cell binding to the Sema7A–Fc-coated surface, indicating that plexin C1 is not a binding partner for Sema7A in monocytes. As previously reported[6], Sema7A–Fc induced the production of pro-inflammatory cytokines in both human peripheral blood monocytes and mouse bone-marrow-derived macrophages (BMDMs) (Supplementary Fig. 1). Sema7A-induced production of interleukin (IL)-6 from monocytes was also significantly inhibited by anti-α1 or anti-β1 monoclonal antibody alone, and the combination of these antibodies resulted in further inhibition (Fig. 1b). These results suggest that Sema7A binds and stimulates monocytes through α1β1 integrin, which is known to be a collagen receptor[10,11]. The RGD sequence, which is frequently found in proteins recognized by integrins[8], is necessary for the activity of Sema7A on neuronal cells[5]. Site-directed mutagenesis of the RGD sequence in Sema7A–Fc or treatment of THP-1 cells with synthetic RGD peptides blocked cell adhesion (Supplementary Fig. 2), suggesting that α1β1 integrin recognizes the RGD sequence of Sema7A. We next tested the ability of Sema7A to stimulate BMDMs from α1-integrin-deficient (*Itga1*$^{-/-}$) mice. On Sema7A–Fc-coated surfaces, cytokine production and adhesion were significantly reduced in *Itga1*$^{-/-}$ BMDMs as compared with wild-type cells (Fig. 1c; see also Supplementary Fig. 3). These observations reinforce that α1β1 integrin is the major receptor for Sema7A.

We further examined the interaction between Sema7A and α1β1 integrin using a soluble form of integrin heterodimer. Treatment of the Sema7A–Fc-coated surface with soluble α1β1 integrin, but not soluble α2β1 or αvβ3 integrin, significantly inhibited adhesion of THP-1 cells (Fig. 1d). We also confirmed binding of Sema7A–Fc to α1β1 integrin by an enzyme-linked immunosorbent assay (ELISA) and a co-immunoprecipitation assay (Supplementary Fig. 4a, b). The interaction between these proteins was inhibited with anti-β1 and/or anti-α1 monoclonal antibodies or RGD peptide (Supplementary Fig. 4c). These observations indicate that Sema7A directly interacts with α1β1 integrin.

Once bound to their ligands, integrins form clusters on the cell surface, leading to the recruitment of signalling and cytoskeletal proteins to form multimolecular signalling modules called focal adhesion complexes[8]. When THP-1 cells were stimulated with Sema7A–Fc, both α1 and β1 integrins were redistributed to the cell periphery, as was focal adhesion kinase (FAK), which is an essential

[1]Department of Molecular Immunology and CREST program of JST, and [2]Department of Immunopathology, Research Institute for Microbial Diseases, Osaka University, 3-1 Yamada-oka Suita, Osaka 565-0871, Japan. [3]Department of Neurology, Graduate School of Medicine, Osaka University, 2-2 Yamada-oka, Suita, Osaka 565-0871, Japan. [4]Department of Pharmacology and Anatomy, Rudolf Magnus Institute of Neuroscience, University Medical Center Utrecht, Universiteitsweg 100, 3584 CG Utrecht, The Netherlands. [5]Laboratory of Protein Synthesis and Expression, Institute for Protein Research, Osaka University, 3-2 Yamada-oka, Suita, Osaka 565-0871, Japan. [6]Biogen-Idec Inc, 12 Cambridge Center, Cambridge, Massachusetts 01746, USA. [7]Howard Hughes Medical Institute and The Solomon H. Snyder Department of Neuroscience, The Johns Hopkins University School of Medicine, 725 North Wolfe Street, Baltimore, Maryland 21205, USA.

component of focal adhesion complexes and a pivotal signal transducer downstream of integrins[12,13] (Fig. 2a). This indicates that Sema7A directly induces clustering of α1β1 integrin and formation of focal adhesion complexes. In neuronal cells, Sema7A activates mitogen-activated protein kinase (MAPK) signalling cascades, which emanate from β1 integrin receptors[5]. Similarly, Sema7A induced phosphorylation of both FAK and extracellular regulated kinases (ERK) 1 and 2 in THP-1 cells (Fig. 2b). Indeed, U0126, a specific inhibitor of MAPK kinase (MEK)-1 and MEK-2 that blocks ERK1/2 phosphorylation, inhibited Sema7A-induced cytokine production in a dose-dependent manner, whereas the analogous but inactive compound U0124 had little effect (Fig. 2c).

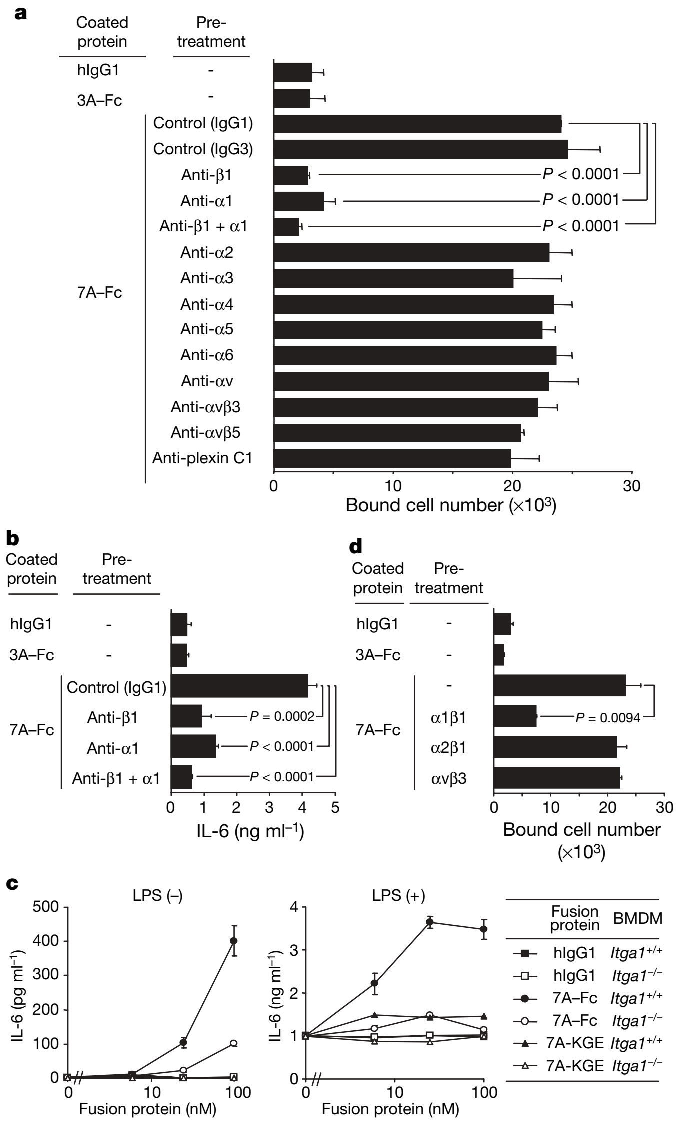


Figure 1 | α1β1 integrin is the functional receptor for Sema7A. **a**, THP-1 cells were treated with the indicated anti-integrin monoclonal antibodies or anti-plexin C1 monoclonal antibody (25 μg ml^{-1}) and incubated on microtitre plates coated with 10 nM human IgG1, Sema3A–Fc or Sema7A–Fc. The number of bound cells was determined by fluorescence-based quantification of nucleic acids. **b**, Human monocytes were treated with the indicated monoclonal antibodies (25 μg ml^{-1}) and cultured on microtitre plates coated with 25 nM fusion proteins. **c**, *Itga1*$^{+/+}$ or *Itga1*$^{-/-}$ BMDMs were cultured on microtitre plates coated with various concentrations of human IgG1, Sema7A–Fc or its RGD-mutated protein (Sema7A-KGE) in the presence or absence of lipopolysaccharide (LPS; 10 ng ml^{-1}). **d**, Sema7A-coated wells were treated with the indicated soluble integrin proteins (50 μg ml^{-1}) before adhesion of THP-1 cells. In **b** and **c**, IL-6 concentration in the culture supernatants was measured by ELISA. Data are the mean ± s.d. of triplicate wells and are representative of at least three independent experiments.

Considering its predominant expression on activated T cells (Supplementary Figs 5 and 6) and its potent stimulation of macrophages, we proposed that Sema7A is involved in macrophage activation by T cells, which is an important step in the inflammatory process. To analyse the localization of Sema7A during T-cell–macrophage interactions, CD4^{+} T cells from OT-II transgenic mice expressing ovalbumin-specific T-cell receptor (TCR), which were stimulated with anti-CD3 and anti-CD28 monoclonal antibodies to induce Sema7A expression, were conjugated with ovalbumin-peptide-pulsed macrophages. Notably, Sema7A clustered at the contact site between T cells and macrophages, characterized by condensation of the lipid raft marker, GM1 glycosphingolipid (Fig. 2d). Accumulation of α1β1 integrin at the contact site was also observed in macrophages (Fig. 2e). In contrast, both Sema7A in T cells and α1β1 integrin in macrophages were evenly distributed on the surface of unconjugated cells (Supplementary Fig. 7). These results indicate that Sema7A and α1β1 integrin are components of the immunological synapse between T cells and macrophages. Even in the conjugates with *Itga1*$^{-/-}$ macrophages, Sema7A in T cells was recruited to the contact site (Fig. 2f). On the other hand, α1β1 integrin did not redistribute to the synapse when *Sema7a*$^{-/-}$ T cells were used (Fig. 2g). These observations demonstrate that Sema7A in T cells, which is recruited to the immunological synapse along with lipid rafts, induces redistribution of α1β1 integrin in macrophages.

We then analysed the role of Sema7A in the immune response using *Sema7a*$^{-/-}$ mice. In these mice, no abnormalities in cell surface phenotype, number or composition of lymphocytes in the spleen and thymus were observed (data not shown). *Sema7a*$^{-/-}$ T cells and *Sema7a*$^{-/-}$ OT-II TCR transgenic T cells showed responses that were indistinguishable from those of wild-type control cells when stimulated by allogeneic dendritic cells and ovalbumin peptides presented by antigen-presenting cells, respectively (Fig. 3a, b). Thus, Sema7A deficiency does not affect the activation of T cells induced by antigenic stimuli.

We next examined the ability of Sema7A on T cells to induce cytokine production in macrophages. To evaluate the effects mediated through direct cellular contact and minimize the influence of soluble T-cell factors[14], T cells were fixed with paraformaldehyde and cultured with BMDMs. BMDMs cultured with activated *Sema7a*$^{-/-}$ T cells produced considerably lower levels of IL-6 and tumour-necrosis factor (TNF)-α than those cultured with wild-type cells (Fig. 3c). In this setting, naive T cells induced much lower levels of cytokine production; levels were indistinguishable between wild-type and *Sema7a*$^{-/-}$ cells (Supplementary Fig. 8a). Notably, BMDMs lacking α1 integrin, a Sema7A receptor, responded poorly to activated wild-type T cells (Supplementary Fig. 8b). These results are indicative of the importance of Sema7A as an effector molecule expressed on activated T cells.

We investigated the role of Sema7A in T-cell-mediated immunity using two different experimental models: contact hypersensitivity (CHS)[15] and experimental autoimmune encephalomyelitis (EAE)[16]. Wild-type mice that had been sensitized with 2,4-dinitrofluorobenzene (DNFB) mounted typical CHS responses characterized by ear swelling and infiltration of mononuclear cells on re-challenge with DNFB. In contrast, *Sema7a*$^{-/-}$ mice were defective in eliciting CHS responses to DNFB (Fig. 3d; see also Supplementary Fig. 9a). Moreover, *Sema7a*$^{-/-}$ mice were highly resistant to EAE induction when immunized with myelin oligodendrocyte glycoprotein (MOG) peptide, and exhibited very few infiltrating cells in the spinal cord (Fig. 3e; see also Supplementary Fig. 9b).

To determine the contribution of Sema7A to T-cell priming in both experimental models, T cells were isolated from draining lymph nodes after epicutaneous DNFB sensitization or MOG immunization, and stimulated with either dinitrobenzene sulphonate (DNBS, a water-soluble analogue of DNFB) or MOG peptide presented by

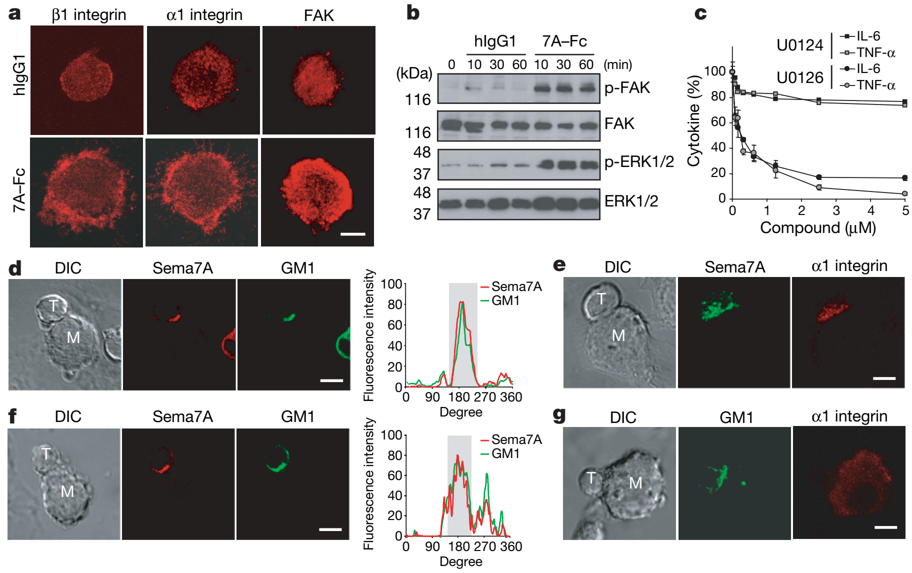


Figure 2 | The mechanisms by which Sema7A stimulates monocytes and macrophages. a, b, Distribution of β1 integrin, α1 integrin and FAK (**a**) and phosphorylation of FAK and ERK1/2 (**b**) were assessed in THP-1 cells that were attached to human IgG1- or Sema7A–Fc-coated surfaces. **c,** Human monocytes were pre-treated with vehicle (0.001% DMSO), various concentrations of U0126 or U0124 (a negative control), and then stimulated on plates coated with 25 nM Sema7A–Fc. Concentrations of the indicated cytokines were determined by ELISA and are expressed as a percentage of control (vehicle) values ± s.d. **d–g,** Activated OT-II TCR transgenic $CD4^+$ T cells (wild type in **d–f**; *Sema7a*$^{-/-}$ in **g**) were mixed with ovalbumin-peptide-pulsed BMDMs (wild type in **d**, **e**, **g**; *Itga1*$^{-/-}$ in **f**) and the cell conjugates were stained for Sema7A (red in **d**, **f**; green in **e**) and/or α1 integrin (red in **e**, **g**). GM1 glycosphingolipid (green in **d**, **f**, **g**) in T cells was labelled with the cholera toxin B subunit before conjugate formation. In **d** and **f**, the fluorescence on the T-cell membrane was plotted as pixel values (*y* axis) relative to the position defined as an angle around the cell surface (*x* axis). The shaded areas correspond to the contact region. DIC, differential interference contrast (microscopy). Scale bars, 5 μm. Data are representative of three independent experiments.

syngeneic splenocytes, respectively. *Sema7a*$^{-/-}$ T cells showed proliferative responses and levels of cytokine production—including interferon-γ and IL-17—that were comparable to those of wild-type T cells (Fig. 4a, b). Thus, antigen-specific priming of T cells and generation of cytokine-producing effector T cells are not impaired in *Sema7a*$^{-/-}$ mice.

To examine whether Sema7A on T cells is required for the effector phase of T-cell-mediated immunity, we transferred T cells from DNFB-sensitized wild-type or *Sema7a*$^{-/-}$ mice into wild-type or *Sema7a*$^{-/-}$ recipients, and then challenged with DNFB. Although wild-type T cells induced CHS at almost the same intensity in the two different recipients, *Sema7a*$^{-/-}$ T cells failed to elicit CHS responses even in wild-type recipients (Fig. 4c). Moreover, mice that received T cells from MOG-immunized *Sema7a*$^{-/-}$ mice developed attenuated EAE. (Fig. 4d; see also Supplementary Fig. 10). Thus, Sema7A on antigen-primed T cells is crucial for the effector phase of T-cell-mediated inflammation.

We performed detailed analyses to characterize the involvement of Sema7A in the effector phase of CHS. We first examined the effects of Sema7A deficiency on the recruitment of effector T cells into inflamed skin. T cells from DNFB-sensitized wild-type or *Sema7a*$^{-/-}$ mice were labelled with 5,6-carboxyfluorescein diacetate, succinimidyl ester (CFSE), and injected intravenously into recipient mice. Both wild-type and *Sema7a*$^{-/-}$ donor cells were effectively recruited into DNFB-treated ears, but not vehicle-treated ears, in wild-type and *Sema7a*$^{-/-}$ recipients (Fig. 4e). This indicates that Sema7A is not necessary for T-cell trafficking to inflamed cutaneous sites.

The normal recruitment of *Sema7a*$^{-/-}$ T cells into inflamed tissue indicates that *Sema7a*$^{-/-}$ T cells might be impaired in the capacity to induce inflammation. To test this possibility, T cells from DNFB-sensitized wild-type or *Sema7a*$^{-/-}$ mice were injected intradermally

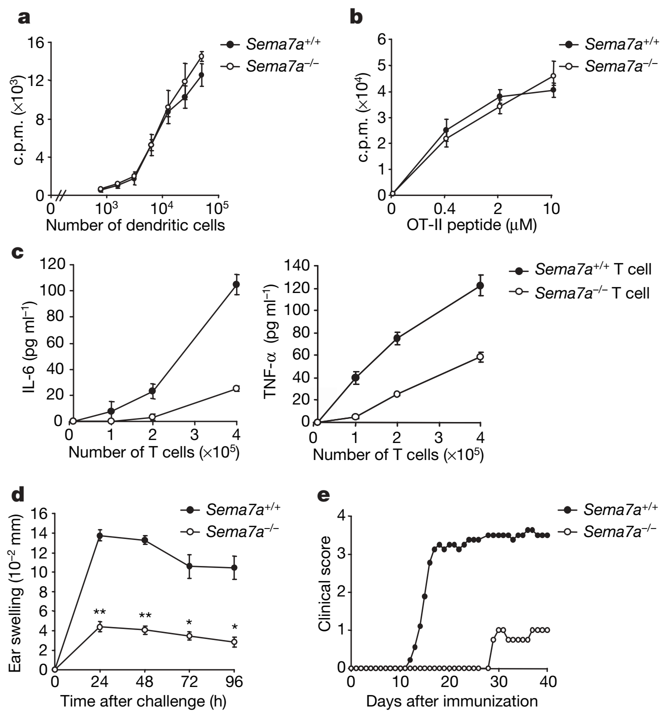


Figure 3 | T-cell function and T-cell-mediated immunity in *Sema7a*$^{-/-}$ mice. a, b, $CD4^+$ T cells (**a**) or OT-II TCR transgenic $CD4^+$ T cells (**b**) from *Sema7a*$^{+/+}$ or *Sema7a*$^{-/-}$ mice were cultured with allogeneic dendritic cells or ovalbumin-peptide-pulsed splenocytes, respectively. Proliferation was assessed by $[^3H]$thymidine uptake. c.p.m., counts per min. **c,** BMDMs were cultured with activated *Sema7a*$^{+/+}$ or *Sema7a*$^{-/-}$ T cells that were fixed with paraformaldehyde. Concentrations of IL-6 and TNF-α in the culture supernatants were measured by ELISA. Data in **a–c** are the mean ± s.d. of triplicate wells. **d,** CHS was induced with DNFB in *Sema7a*$^{+/+}$ or *Sema7a*$^{-/-}$ mice ($n = 5$ per group). Net ear swelling is shown as the mean ± s.e.m. Asterisk, $P < 0.01$; double asterisk, $P < 0.001$. **e,** *Sema7a*$^{+/+}$ ($n = 8$) or *Sema7a*$^{-/-}$ ($n = 7$) mice were immunized with 100 μg of MOG(35–55) peptide and scored for clinical signs of EAE as described[27]. Data are representative of at least three independent experiments.

into the ear tissue of wild-type recipients, followed by DNFB challenge. Notably, the intensity of the ear-swelling response induced by sensitized *Sema7a*$^{-/-}$ T cells was markedly attenuated as compared with that induced by wild-type cells (Fig. 4f). Thus, Sema7A on antigen-primed T cells is required for the optimal effector function of these cells in enhancing inflammation at the site of CHS.

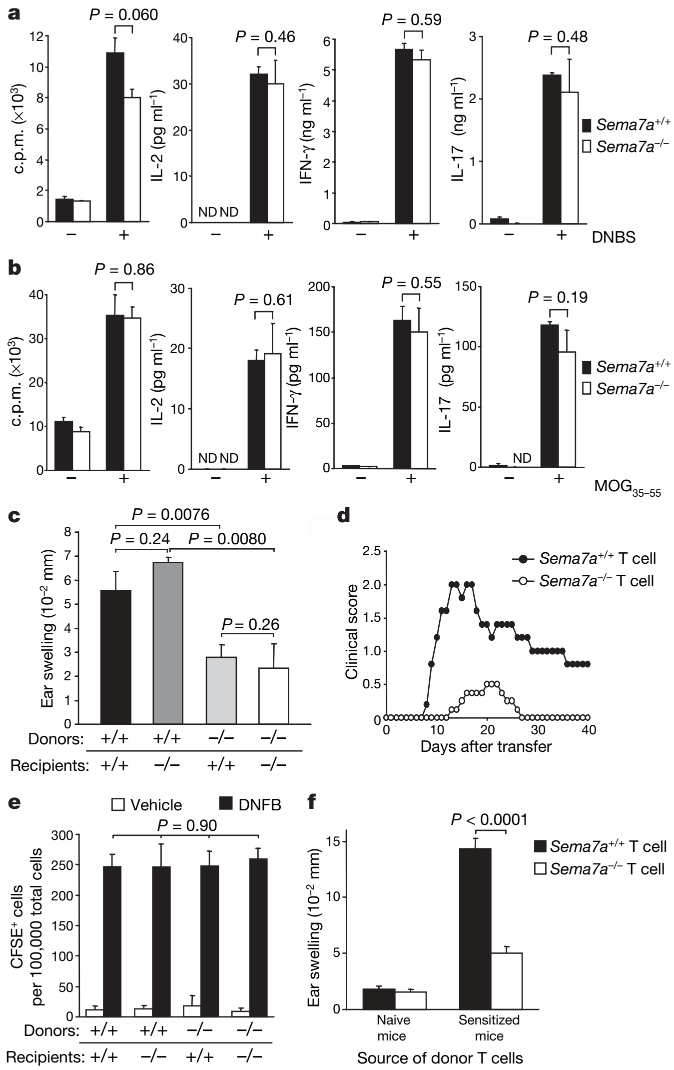


Figure 4 | Involvement of Sema7A in the effector phase of the immune response. **a**, **b**, Thy1.2^{+} cells from DNFB-sensitized mice (**a**) or CD4^{+} cells from MOG-immunized mice (**b**) were re-stimulated with the corresponding antigen. Proliferation was assessed by [^{3}H]thymidine uptake, and cytokines in the culture supernatants were measured using Bio-Plex suspension array systems. ND, not detected. **c**, *Sema7a*$^{+/+}$ (+/+) or *Sema7a*$^{-/-}$ (−/−) mice ($n = 5$ per group) intravenously received DNFB-primed *Sema7a*$^{+/+}$ or *Sema7a*$^{-/-}$ T cells, followed by DNFB challenge. Ear swelling responses were measured 48 h later. **d**, MOG-stimulated CD4^{+} T cells were transferred into sub-lethally irradiated mice ($n = 7$ for *Sema7a*$^{+/+}$ T cells; $n = 8$ for *Sema7a*$^{-/-}$ T cells), and clinical signs of EAE were scored daily. **e**, DNFB-primed *Sema7a*$^{+/+}$ (+/+) or *Sema7a*$^{-/-}$ (−/−) T cells were labelled with CFSE and injected intravenously into *Sema7a*$^{+/+}$ or *Sema7a*$^{-/-}$ mice ($n = 4$ per group). Single cell suspensions were prepared from vehicle- or DNFB-treated ears 12 h after transfer and analysed by flow cytometry. The mean number of CFSE^{+} cells per 100,000 total cells in the ear is shown. The statistical analysis was performed using analysis of variance. **f**, T cells from naive or DNFB-sensitized *Sema7a*$^{+/+}$ or *Sema7a*$^{-/-}$ mice were injected intradermally into the ear of recipients ($n = 5$ per group). CHS responses were determined 48 h after DNFB challenge. Data are the mean ± s.d. (**a**, **b**) or mean ± s.e.m. (**c**, **e**, **f**), and are representative of two (**c**, **e**, **f**) or three (**a**, **b**, **d**) independent experiments.

The present study demonstrates that Sema7A has a critical role in T-cell-mediated inflammation through α1β1 integrin on macrophages. In line with our results, the reported phenotypes of *Itga1*$^{-/-}$ mice and mice treated with anti-α1 integrin monoclonal antibody are similar to those of *Sema7a*$^{-/-}$ mice[17–19]. Previous studies have emphasized the interactions between α1β1 integrin on effector T cells and extracellular matrix proteins (such as collagen and laminin), with particular attention given to the roles of extracellular matrix proteins in the retention and survival of T cells at inflammatory sites[10,20,21]. In several animal models of inflammation, however, α1β1 integrin is predominantly expressed by monocytes and macrophages[17–19]. Indeed, *Itga1*$^{-/-}$ mice that received encephalitogenic T cells developed less severe EAE as compared with wild-type recipients (Supplementary Fig. 11), which further supports the importance of the interaction between Sema7A on T cells and α1β1 integrin on monocytes and macrophages. We have shown that both Sema7A and α1β1 integrin are clustered at the immunological synapses between T cells and macrophages. This may enable Sema7A to induce clustering and activation of integrins more efficiently than extracellular matrix proteins, which contributes to initiating the effector cell programme of T-cell-mediated inflammation (Supplementary Fig. 13).

We cannot, however, exclude a possible involvement of other receptors in Sema7A signalling. Indeed, *Itga1*$^{-/-}$ macrophages still produced small amounts of cytokine when stimulated with Sema7A–Fc (Fig. 1c). Furthermore, cytokine production by *Itga1*$^{-/-}$ macrophages induced by activated *Sema7a*$^{-/-}$ T cells decreased slightly compared with production induced by wild-type T cells (Supplementary Fig. 8b). Thus, although Sema7A signalling is largely mediated by α1β1 integrin, a small contribution might be made by other molecules, possibly β1-containing integrins, that we have not tested.

Although it has been reported that Sema7A is a negative regulator of T-cell activation[7], we did not observe hyperactivation of *Sema7a*$^{-/-}$ T cells or hypersensitivity of *Sema7a*$^{-/-}$ mice in cellular immune responses. In the previous study[7], *Sema7a*$^{-/-}$ mice developed highly severe EAE when immunized with extremely low doses of MOG peptide. However, we did not observe exacerbated EAE in *Sema7a*$^{-/-}$ mice at any dose of MOG peptide during the course of our study, including low doses of antigen similar to those used in ref. 7 (Supplementary Fig. 12). The previous report also suggested that Sema7A negatively regulates T-cell-receptor signalling at the initial phase of antigen-induced T-cell activation. However, the absence of Sema7A on naive T cells might not allow for this model. Therefore, we do not know to what extent, if any, an inhibitory role of Sema7A contributes to physiological or pathological immune responses.

Our findings have established that integrin-mediated signalling is a common mechanism by which Sema7A functions in both the nervous and immune systems, and that Sema7A has an essential function in the effector phase of cell-mediated immunity. Clinical symptoms of autoimmune and chronic inflammatory diseases are direct manifestations of the effector phases of aberrant inflammatory reactions. Therefore, the interaction between Sema7A and α1β1 integrin is a potential therapeutic target for these immune disorders.

METHODS

Full Methods are available in Supplementary Information, unless described below.

Mice. C57BL/6 and BALB/c mice were purchased from Clea. *Sema7a*$^{-/-}$ (ref. 5) and *Itga1*$^{-/-}$ (ref. 22) mice on the C57BL/6 background were generated as previously described. OT-II TCR transgenic mice were provided by W. R. Heath[23]. OT-II TCR transgenic mice on the *Sema7a*$^{-/-}$ background were obtained by crossing OT-II TCR transgenic mice with *Sema7a*$^{-/-}$ mice. Six- to eight-week-old mice were used for experiments. All mice used in this study were housed in specific pathogen-free conditions. Animal experimentation protocols were reviewed and approved by our institutional animal care committees.

Cell culture. Human monocytes were isolated from the peripheral blood of volunteers using the RosseteSep human monocyte enrichment cocktail

(StemCell Technologies). Mouse BMDMs were obtained from 5-day cultures of bone marrow cells in the presence of 50 ng ml^{-1} macrophage colony-stimulating factor (Peprotec) as previously described[24]. Cells (1×10^5 per well) were added to flat-bottomed 96-well plates coated with the indicated concentrations of human IgG1 (Calbiochem), Sema3A–Fc (R&D Systems) or Sema7A–Fc and cultured for 24 h in RPMI1640 medium containing 10% FCS. For T-cell–BMDM co-cultures[14], Thy1.2$^+$ cells were purified from mouse spleens with an automated magnetic cell sorter (AutoMACS, Miltenyi Biotech) and stimulated with immobilized anti-CD3 (2C11, 10 μg ml^{-1}, BD Pharmingen) and anti-CD28 (37.51, 5 μg ml^{-1}, BD Pharmingen) monoclonal antibodies. Three days later, cells were collected and fixed with 1% paraformaldehyde in PBS for 2 h at 4 °C. The indicated numbers of fixed cells were cultured with BMDMs (1×10^5 cells per well) in flat-bottomed 96-well plates for 24 h. Cytokine concentration in the culture supernatants was measured with ELISA kits (R&D Systems). T-cell proliferation assays were performed as previously described[25].

Cell adhesion assay. The procedure for the cell adhesion assay was carried out as previously described[26]. The number of attached cells was determined with a CyQUANT cell proliferation assay kit (Molecular Probes). For antibody blocking assays, cells were pre-treated with 25 μg ml^{-1} anti-integrin monoclonal antibodies or anti-human plexin C1 monoclonal antibody (M460, provided by Amgen) for 30 min on ice. Recombinant soluble α1β1, α2β1 and αvβ3 integrin proteins were used for treatment of Sema7A–Fc-coated wells at 50 μg ml^{-1} and plates were incubated for 3 h at room temperature before cell adhesion.

Statistical analysis. To analyse statistical significance, we used an unpaired, two-tailed Student's *t*-test, unless specified otherwise. We considered P-values <0.05 to be significant.

Received 25 December 2006; accepted 5 February 2007.
Published online 21 March 2007.

1. Kolodkin, A. L., Matthes, D. J. & Goodman, C. S. The semaphorin genes encode a family of transmembrane and secreted growth cone guidance molecules. *Cell* **75**, 1389–1399 (1993).
2. Kikutani, H. & Kumanogoh, A. Semaphorins in interactions between T cells and antigen-presenting cells. *Nature Rev. Immunol.* **3**, 159–167 (2003).
3. Lange, C. *et al.* New eukaryotic semaphorins with close homology to semaphorins of DNA viruses. *Genomics* **51**, 340–350 (1998).
4. Xu, X. *et al.* Human semaphorin K1 is glycosylphosphatidylinositol-linked and defines a new subfamily of viral-related semaphorins. *J. Biol. Chem.* **273**, 22428–22434 (1998).
5. Pasterkamp, R. J., Peschon, J. J., Spriggs, M. K. & Kolodkin, A. L. Semaphorin 7A promotes axon outgrowth through integrins and MAPKs. *Nature* **424**, 398–405 (2003).
6. Holmes, S. *et al.* Sema7A is a potent monocyte stimulator. *Scand. J. Immunol.* **56**, 270–275 (2002).
7. Czopik, A. K., Bynoe, M. S., Palm, N., Raine, C. S. & Medzhitov, R. Semaphorin 7A is a negative regulator of T cell responses. *Immunity* **24**, 591–600 (2006).
8. Hynes, R. O. Integrins: bidirectional, allosteric signaling machines. *Cell* **110**, 673–687 (2002).
9. Tamagnone, L. *et al.* Plexins are a large family of receptors for transmembrane, secreted, and GPI-anchored semaphorins in vertebrates. *Cell* **99**, 71–80 (1999).
10. Ben-Horin, S. & Bank, I. The role of very late antigen-1 in immune-mediated inflammation. *Clin. Immunol.* **113**, 119–129 (2004).
11. Hemler, M. E. VLA proteins in the integrin family: structures, functions, and their role on leukocytes. *Annu. Rev. Immunol.* **8**, 365–400 (1990).
12. Mitra, S. K., Hanson, D. A. & Schlaepfer, D. D. Focal adhesion kinase: in command and control of cell motility. *Nature Rev. Mol. Cell Biol.* **6**, 56–68 (2005).
13. Schlaepfer, D. D. & Hunter, T. Integrin signalling and tyrosine phosphorylation: just the FAKs? *Trends Cell Biol.* **8**, 151–157 (1998).
14. McInnes, I. B., Leung, B. P., Sturrock, R. D., Field, M. & Liew, F. Y. Interleukin-15 mediates T cell-dependent regulation of tumor necrosis factor-α production in rheumatoid arthritis. *Nature Med.* **3**, 189–195 (1997).
15. Grabbe, S. & Schwarz, T. Immunoregulatory mechanisms involved in elicitation of allergic contact hypersensitivity. *Immunol. Today* **19**, 37–44 (1998).
16. Kuchroo, V. K. *et al.* T cell response in experimental autoimmune encephalomyelitis (EAE): role of self and cross-reactive antigens in shaping, tuning, and regulating the autopathogenic T cell repertoire. *Annu. Rev. Immunol.* **20**, 101–123 (2002).
17. de Fougerolles, A. R. *et al.* Regulation of inflammation by collagen-binding integrins α1β1 and α2β1 in models of hypersensitivity and arthritis. *J. Clin. Invest.* **105**, 721–729 (2000).
18. Fiorucci, S. *et al.* Importance of innate immunity and collagen binding integrin α1β1 in TNBS-induced colitis. *Immunity* **17**, 769–780 (2002).
19. Krieglstein, C. F. *et al.* Collagen-binding integrin α1β1 regulates intestinal inflammation in experimental colitis. *J. Clin. Invest.* **110**, 1773–1782 (2002).
20. Andreasen, S. O. *et al.* Expression and functional importance of collagen-binding integrins, α1β1 and α2β1, on virus-activated T cells. *J. Immunol.* **171**, 2804–2811 (2003).
21. Ray, S. J. *et al.* The collagen binding α1β1 integrin VLA-1 regulates CD8 T cell-mediated immune protection against heterologous influenza infection. *Immunity* **20**, 167–179 (2004).
22. Gardner, H., Kreidberg, J., Koteliansky, V. & Jaenisch, R. Deletion of integrin α1 by homologous recombination permits normal murine development but gives rise to a specific deficit in cell adhesion. *Dev. Biol.* **175**, 301–313 (1996).
23. Barnden, M. J., Allison, J., Heath, W. R. & Carbone, F. R. Defective TCR expression in transgenic mice constructed using cDNA-based α- and β-chain genes under the control of heterologous regulatory elements. *Immunol. Cell Biol.* **76**, 34–40 (1998).
24. Riches, D. W. & Underwood, G. A. Expression of interferon-β during the triggering phase of macrophage cytocidal activation. Evidence for an autocrine/paracrine role in the regulation of this state. *J. Biol. Chem.* **266**, 24785–24792 (1991).
25. Kumanogoh, A. *et al.* Requirement for the lymphocyte semaphorin, CD100, in the induction of antigen-specific T cells and the maturation of dendritic cells. *J. Immunol.* **169**, 1175–1181 (2002).
26. Hanayama, R. *et al.* Identification of a factor that links apoptotic cells to phagocytes. *Nature* **417**, 182–187 (2002).
27. Riminton, D. S. *et al.* Challenging cytokine redundancy: inflammatory cell movement and clinical course of experimental autoimmune encephalomyelitis are normal in lymphotoxin-deficient, but not tumor necrosis factor-deficient, mice. *J. Exp. Med.* **187**, 1517–1528 (1998).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature. A summary figure is also included.

Acknowledgements We thank K. Kubota for secretarial assistance. We are grateful to R. Hanayama, T. Toyofuku, T. Yasui and J. Encinas for critical advice, discussion and encouragement. We also thank T. Sugimoto, K. Shiozaki, I. Sizing and G. Majeau for technical support. This study was supported by the following funding agencies: the Ministry of Education, Culture, Sports, Science and Technology, Japan and CREST programme of JST (A.K. and H.K.); the Program for Promotion of Fundamental Studies in Health Sciences of the National Institute of Biomedical Innovation (A.K.); NINDS/NIH (A.L.K.); The Netherlands Organization for Scientific Research, The Human Frontier Science Program Organization and National Alliance for Research on Schizophrenia and Depression (R.J.P.); and Research Fellowships of the Japan Society for the Promotion of Science for Young Scientists (K.S. and M.Y.). A.L.K. is an investigator of the Howard Hughes Medical Institute.

Author Contributions K.S. performed the main experimental work, analysed the results and wrote the manuscript. T.O., M.Y., N.T., H.T. and T.K. also performed experimental work. R.J.P. and A.L.K. produced *Sema7a*$^{-/-}$ mice. J.T. produced recombinant integrins. P.D.R. was involved in the study using *Itga1*$^{-/-}$ mice. A.K. and H.K. co-organized and performed project planning, data analysis and writing the manuscript

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.K. (kumanogo@ragtime.biken.osaka-u.ac.jp) or H.K. (kikutani@ragtime.biken.osaka-u.ac.jp).

Vol 446 | 5 April 2007 | doi:10.1038/nature05673

Foxp3 controls regulatory T-cell function by interacting with AML1/Runx1

Masahiro Ono[1,2]*, Hiroko Yaguchi[3]*, Naganari Ohkura[3]*, Issay Kitabayashi[4], Yuko Nagamura[3], Takashi Nomura[1], Yoshiki Miyachi[2], Toshihiko Tsukada[3] & Shimon Sakaguchi[1,5]

Naturally arising $CD25^{+}CD4^{+}$ regulatory T cells (T_R cells) are engaged in the maintenance of immunological self-tolerance and immune homeostasis by suppressing aberrant or excessive immune responses, such as autoimmune disease and allergy[1–3]. T_R cells specifically express the transcription factor Foxp3, a key regulator of T_R-cell development and function. Ectopic expression of Foxp3 in conventional T cells is indeed sufficient to confer suppressive activity, repress the production of cytokines such as interleukin-2 (IL-2) and interferon-gamma (IFN-γ), and upregulate T_R-cell-associated molecules such as CD25, cytotoxic T-lymphocyte-associated antigen-4, and glucocorticoid-induced TNF-receptor-family-related protein[4–7]. However, the method by which Foxp3 controls these molecular events has yet to be explained. Here we show that the transcription factor AML1 (acute myeloid leukaemia 1)/Runx1 (Runt-related transcription factor 1), which is crucially required for normal haematopoiesis including thymic T-cell development[8–11], activates *IL-2* and *IFN-γ* gene expression in conventional $CD4^{+}$ T cells through binding to their respective promoters. In natural T_R cells, Foxp3 interacts physically with AML1. Several lines of evidence support a model in which the interaction suppresses IL-2 and IFN-γ production, upregulates T_R-cell-associated molecules, and exerts suppressive activity. This transcriptional control of T_R-cell function by an interaction between Foxp3 and AML1 can be exploited to control physiological and pathological T-cell-mediated immune responses.

Functional studies of the *IL-2* promoter have ascertained that a minimal promoter region extending to 300 base pairs (bp) relative to the transcription start site of the *IL-2* gene, containing binding sites for nuclear factor of activated T cells (NFAT), Oct1, activator protein-1 (AP-1) and nuclear factor-κB (NF-κB)[12], is sufficient for inducible expression of the gene after stimulation of the T-cell antigen receptor (TCR) in reporter assays[13]. However, the upstream sequences beyond this region also enhance promoter activity[12]. By examining the 2.0-kb 5′ flanking region of the mouse *IL-2* gene, we have found three putative AML1–DNA binding consensus sites (5′-ACCACA-3′) at −370, −1,327 and −1,458 bp (relative to the transcription start site), denoted RE1, RE2 and RE3, respectively (Fig. 1a). In particular, a region containing the proximal RE1 site is highly conserved between humans, mice and other mammalian species (Supplementary Fig. S1) and is reported to be selectively and rapidly demethylated on T-cell activation[14].

To determine whether AML1 binds to the *IL-2* promoter and modifies *IL-2* gene expression, we performed several assays with mouse primary $CD4^{+}$ T cells and human Jurkat cells. In chromatin immunoprecipitation (ChIP) assays using an antibody against AML1, the *IL-2* promoter region adjacent to the RE1 site was co-precipitated in mouse primary $CD4^{+}$ T cells, whereas the 5′ and 3′ regions far from the start site were not (Fig. 1b). The human *IL-2* promoter region encompassing the AML1 site corresponding to the mouse RE1 site was also co-precipitated with the anti-AML1 antibody in Jurkat cells (Fig. 1b). Electrophoretic mobility-shift assays showed specific retardation of the complexes formed between the DNA-binding domain of AML1 and radiolabelled oligonucleotides corresponding to RE1, RE2 and RE3 (Fig. 1c). Moreover, mutating the RE1, RE2 and RE3 sites by altering two nucleotides at each site (mut RE1, mut RE2 and mut RE3; Supplementary Fig. S2) completely abolished AML1 binding (Fig. 1c).

To examine the functional responsiveness of the *IL-2* promoter to AML1, we made a reporter construct composed of the 1.6-kb *IL-2* promoter, containing the three AML1-binding sites, fused to a luciferase reporter gene. Stimulation of Jurkat cells with anti-CD3 and anti-CD28 activated the *IL-2* promoter in the construct[15], and transfection of AML1 enhanced this activation further (Fig. 1d). The use of a mutated *IL-2* promoter with mut RE1, mut RE2 and mut RE3 sites (*IL-2* promoter ΔAML-*luc*) resulted in a striking decrease in both stimulation-induced and AML1-enhanced promoter activities (Fig. 1e). Mutations that disrupted NFAT sites (*IL-2* promoter ΔNFAT-*luc*) led to a complete elimination of both stimulation-induced transcription activation and AML1-enhanced transactivation (Fig. 1f), suggesting that AML1-dependent activation of the *IL-2* promoter is reliant on the NFAT transcription factor complex.

Knockdown of endogenous *AML1* by RNA interference (RNAi) in Jurkat cells resulted in a marked decrease in IL-2 production after stimulation, indicating that AML1 is physiologically required for the induction of IL-2 expression in activated T cells (Fig. 1g). Moreover, in mouse primary T cells retrovirally transduced with *AML1*, IL-2 production in response to anti-CD3 and anti-CD28 stimulation was much higher than in empty-vector-transduced T cells (Fig. 1h, i). Conversely, IL-2 production was markedly reduced in mouse T cells transduced with a dominant-negative form of AML1 (AML1-DN, also called AML1a (ref. 16); see Fig. 2d), which lacks the C-terminal transcriptional activation and inhibition domains (Fig. 1h, i).

Taken together, the findings in Fig. 1 indicate that AML1 binds specifically to the AML1 sites present in the *IL-2* promoter region, thus enhancing *IL-2* gene expression.

Expression of AML1 protein in both conventional T cells and T_R cells (Supplementary Fig. S3) and strong repression of IL-2 expression by Foxp3 (Fig. 1h, i, refs 4–6) indicate that AML1 and Foxp3 may interact in T_R cells and may thereby control IL-2 expression. Indeed, an anti-FOXP3 antibody co-precipitated endogenous FOXP3 with AML1 from human peripheral blood mononuclear cells (PBMCs) (Fig. 2a). This indicates *in vivo* physiological interaction of

[1]Department of Experimental Pathology, Institute for Frontier Medical Sciences, and [2]Department of Dermatology, Graduate School of Medicine, Kyoto University, Kyoto 606-8507, Japan. [3]Tumor Endocrinology Project, and [4]Molecular Oncology Division, National Cancer Center Research Institute, Chuo-ku Tokyo, 104-0045, Japan. [5]Core Research for Evolutional Science and Technology (CREST), Japan Science and Technology Agency, Kawaguchi 332-0012, Japan.
*These authors contributed equally to this work.

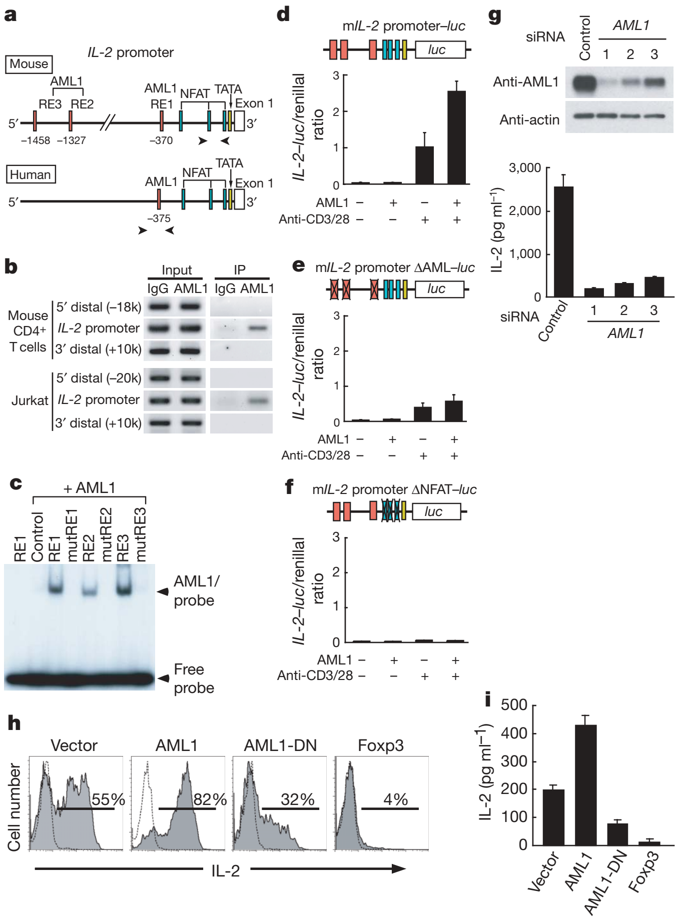


Figure 1 | AML1/Runx1 enhances IL-2 expression through binding to the *IL-2* promoter in activated T cells. a, Schematic representation of the mouse and human *IL-2* promoters. The putative AML1 binding sites, the NFAT sites and the TATA boxes are shown in red, blue and yellow, respectively. **b**, In ChIP assays performed with an antibody against AML1 or control rabbit IgG, aliquots of chromatin obtained before (Input) or after immunoprecipitation (IP) from mouse CD4$^+$ T cells or Jurkat cells were analysed by polymerase chain reaction with primers specific for the *IL-2* promoter. The positions of the ChIP primers are delineated in **a** by arrowheads. The primers for the 5′ and 3′ far distal regions were used as controls. **c**, In an electrophoretic mobility-shift assay, the DNA-binding domain of AML1 was incubated with radiolabelled oligonucleotides spanning the individual AML1 wild-type or mutated (mut) RE1, RE2 and RE3 sites on the *IL-2* promoter (sequences of the oligonucleotides are shown in Supplementary Fig. S2). **d**–**f**, Jurkat cells were simultaneously transfected with either an AML1-expressing or an empty vector, and with wild-type mouse *IL-2* (m*IL-2*) promoter–luciferase (*luc*) (**d**), a mutated promoter ΔAML–*luc* construct containing mut RE1, mut RE2 and mut RE3 (**e**), or a mutated promoter ΔNFAT–*luc* construct containing mutated NFAT consensus sites (**f**). Cells were either unstimulated or stimulated by plate-bound anti-CD3 and soluble anti-CD28 for 6 h, and luciferase activities were measured. Data shown are relative values of firefly luciferase normalized to *Renilla* luciferase and are expressed as means ± s.e.m. **g**, Jurkat cells were transfected with *AML1* siRNAs (1, 2 and 3) or control siRNA, and the expression of endogenous AML1 was evaluated by western blot analysis (top). Cells transfected with the indicated siRNAs were stimulated for 24 h with plate-bound anti-CD3 and soluble anti-CD28, and IL-2 levels in the supernatant were quantified by ELISA (bottom; means ± s.d.). **h**, Mouse CD25$^-$CD4$^+$ T cells were transduced with retroviral pMCsIg vectors encoding indicated proteins, IRES and GFP. GFP-positive cells were sorted and stimulated by plate-bound anti-CD3 and soluble anti-CD28 for 6 h and analysed for intracellular IL-2 expression by flow cytometry. Thin lines represent control staining with an isotype-matched antibody. **i**, Sorted GFP-positive cells were stimulated by soluble anti-CD3 and anti-CD28 with antigen-presenting cells for 24 h, and IL-2 levels in the supernatant were quantified by ELISA (means ± s.d.). Results represent three independent experiments.

endogenous FOXP3 and AML1 in natural T_R cells, because most FOXP3$^+$ T cells in PBMCs are naturally arising CD4$^+$ T_R cells[17–20]. Glutathione *S*-transferase (GST) pulldown experiments showed physical interaction of these two molecules *in vitro*: [^{35}S]methionine-labelled Foxp3 bound to the GST–AML1 fusion protein but not to the GST protein alone, and, reciprocally, labelled AML1 bound to GST–Foxp3 (Fig. 2b). In addition, under confocal microscopy, immunostained endogenous AML1 and Foxp3 were heterogeneously distributed in the nucleus and partly co-localized in mouse CD25$^+$CD4$^+$ T_R cells, suggesting that both proteins comprise part of the same molecular complex in natural T_R cells (Fig. 2c).

We next attempted to determine the Foxp3-interacting domain of AML1 and the AML1-interacting domain of Foxp3. In reciprocal co-immunoprecipitation experiments using lysates of human embryonic kidney 293T cells co-transfected with constructs encoding tagged AML1 and Foxp3 proteins, deletion mutants of AML1 lacking

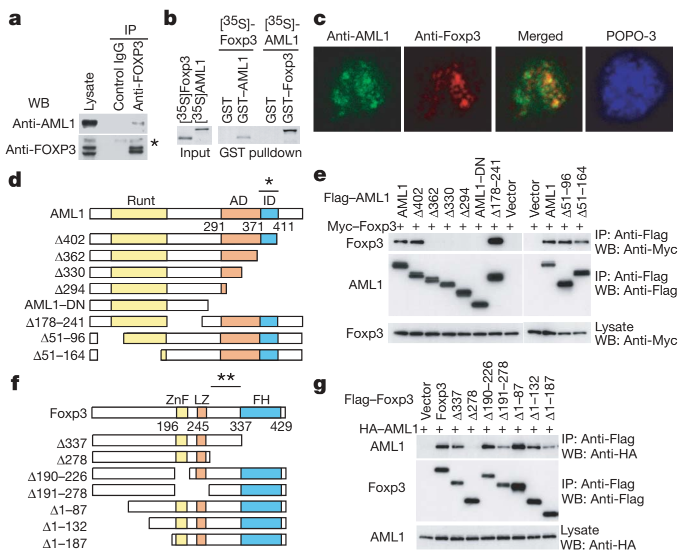


Figure 2 | Foxp3 interacts physically with AML1. a, Endogenous interaction between AML1 and FOXP3 in human PBMCs. PBMCs were lysed, subjected to immunoprecipitation (IP) with anti-FOXP3 antibody or control IgG, and western blotted (WB) with anti-AML1 and anti-FOXP3 antibodies. Asterisk indicates immunoglobulin heavy chain. **b**, *In vitro* interaction between Foxp3 and AML1 was analysed by GST pulldown assays. *In vitro* transcribed/translated [^{35}S]methionine-labelled FoxP3 or AML1 was incubated with bacterially expressed GST–AML1 or GST–Foxp3, respectively. GST alone was included as a negative control. Input consisted of 10% of the [^{35}S]methionine-labelled products. **c**, The intracellular localizations of endogenous AML1 and Foxp3 in mouse T_R cells were analysed by confocal microscopy with antibodies directed against AML1 and Foxp3. A representative image of a single T_R cell is shown. Images of AML1 (green) and Foxp3 (red) were merged to show regions of co-localization (yellow). The nucleus was revealed with POPO-3 staining (blue). **d**, Schematic diagram of wild-type AML1 (shown on top) and the deletion constructs. The Runt, activation and inhibition domains are indicated as Runt, AD and ID, respectively. **e**, Myc-tagged Foxp3 was co-transfected into 293T cells with the indicated Flag-tagged AML1 constructs and immunoprecipitated with anti-Flag M2 agarose; protein blots were probed with anti-Myc (top) or anti-Flag (middle) antibodies. The expression of Foxp3 in the lysates was monitored by immunoblotting with an anti-Myc antibody (bottom). **f**, Schematic diagram of wild-type Foxp3 (top) and the deletion constructs. The zinc finger, leucine zipper and forkhead domains are indicated as ZnF, LZ and FH, respectively. **g**, Haemagglutinin (HA)-tagged AML1 was co-transfected into 293T cells with the indicated Flag-tagged Foxp3 constructs and immunoprecipitated with anti-Flag M2 agarose, and protein blots were probed with anti-HA (top) or anti-Flag (middle) antibodies. The expression of AML1 in the lysates was monitored by immunoblotting with an anti-HA antibody (bottom). Figures represent three independent experiments.

the carboxy-terminal region (amino acids 362–402; asterisk in Fig. 2d) failed to bind to Foxp3, indicating a requirement of this region for interaction with Foxp3 (Fig. 2d, e). This region of AML1 corresponds to the domain that was shown to have inhibitory activity on transcription[21]. Similar experiments using deletion mutants of Foxp3 indicated that the AML1-interacting domain of the Foxp3 protein was located between the forkhead domain and the leucine zipper motif (amino acids 278–336; asterisks in Fig. 2f) (Fig. 2f, g).

ChIP assays showed that Foxp3 bound to the *IL-2* promoter (Supplementary Fig. S4), as reported recently[22]. To establish the functional significance of the interaction between Foxp3 and AML1, we used the same *IL-2* reporter construct described in Fig. 1d to examine whether Foxp3 affects the transactivation activity of AML1 in Jurkat cells (Fig. 3a). Foxp3 did indeed repress both stimulation-induced transcriptional activation and AML1-dependent transcriptional enhancement of the *IL-2* promoter. When the mutated *IL-2* promoter ΔAML1-*luc* (depicted in Fig. 1e) was employed, *IL-2* promoter activity was substantially attenuated and Foxp3 failed to repress the transcription further, suggesting that the repression of *IL-2* transcription by Foxp3 is dependent on AML1 (Fig. 3b).

Next, to determine whether Foxp3 is also able to repress AML1-mediated IL-2 production in mouse primary T cells, we co-transduced $CD25^-CD4^+$ conventional T cells with retroviral constructs of Foxp3–internal ribosome entry site (IRES)–nerve growth factor receptor (NGFR) and AML1–IRES–green fluorescent protein (GFP), and analysed IL-2 production in those T cells co-expressing NGFR and GFP (Fig. 3c, d). Consistent with the results of the luciferase reporter assays was our observation that Foxp3 repressed AML1-driven IL-2 production in the co-transduced T cells (Fig. 3c, d, and Supplementary Fig. S5a). In addition, retroviral gene transduction of AML1 failed to elicit IL-2 production in Foxp3-expressing natural $CD25^+CD4^+$ T_R cells (Fig. 3e, and Supplementary Fig. S5b).

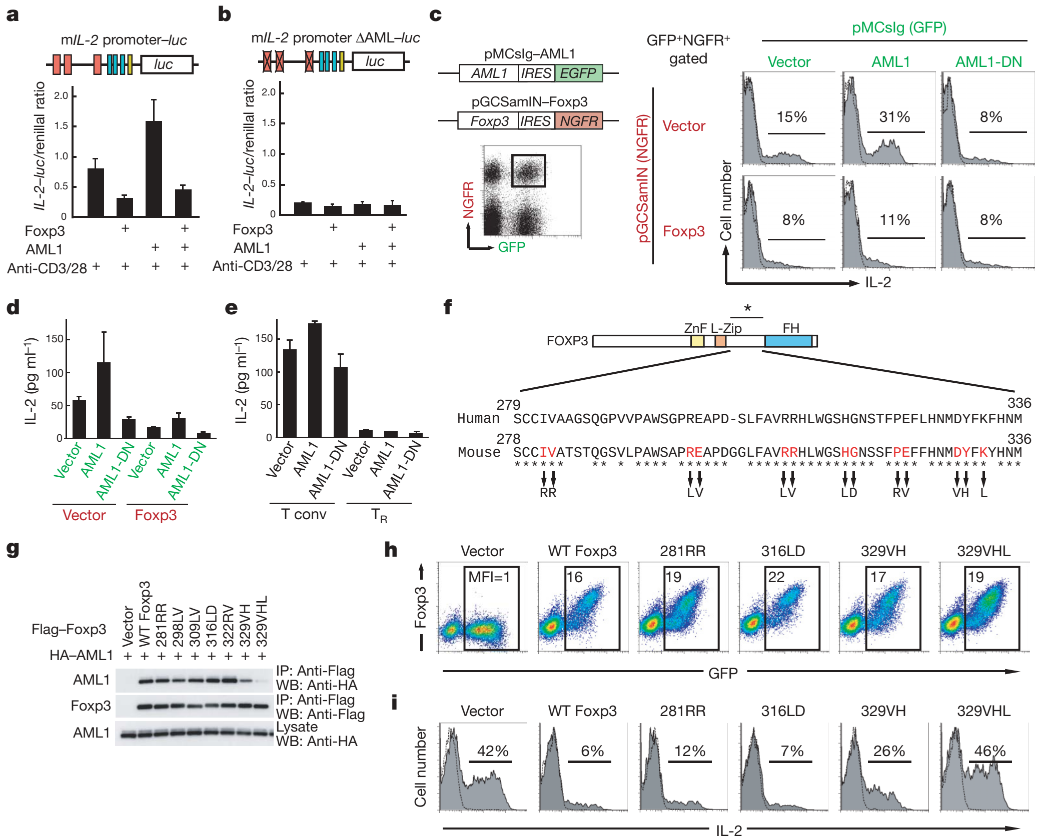


Figure 3 | Foxp3 represses AML1-induced IL-2 expression by interacting with AML1. a, b, Jurkat cells were co-transfected with expression vectors for the indicated proteins and wild-type *IL-2* promoter (**a**) or mutated *IL-2* promoter ΔAML–*luc* constructs (**b**), stimulated as in Fig. 1d–f, and luciferase activities were measured. **c**, Schematic representation of the retroviral vectors, including pMCsIg–AML1 (*AML1–IRES–GFP*) and pGCSamIN–Foxp3 (*Foxp3–IRES–NGFR*), which were used to express AML1 and Foxp3, respectively, in this study. Mouse $CD4^+$ T cells were co-transduced with the pMCsIg and pGCSamIN vectors to express the indicated proteins and were activated by soluble anti-CD3 and anti-CD28 for 6 h. Cells were stained with a biotinylated anti-NGFR antibody and allophycocyanin-conjugated streptavidin, and subsequently fixed and stained with phycoerythrin-conjugated anti-IL-2, and analysed by flow cytometry for intracellular IL-2 in $GFP^+/NFGR^+$ double-positive cells (rectangles). **d**, Co-transduced mouse $CD4^+$ T cells sorted by gating $GFP^+/NFGR^+$ double-positive cells were stimulated as in Fig. 1i, and IL-2 levels in the supernatant were quantified by ELISA. **e**, Mouse $CD25^-CD4^+$ conventional T cells (T conv) or $CD25^+CD4^+$ T cells (T_R) were transduced with the pMCsIg vectors to express the indicated proteins, sorted by gating GFP^+ cells and stimulated as in Fig. 1i; IL-2 levels in the supernatant were quantified by ELISA. Expression of transduced genes in **d** and **e** was confirmed by western blot analysis with sorted GFP^+NGFR^+ (or GFP^+) cells (Supplementary Fig. S5). **f**, Sequence alignment of the AML1-binding region in human and mouse FOXP3. Substitutions were introduced at conserved residues of human/mouse FOXP3 (asterisks) as indicated. **g**, Haemagglutinin (HA)-tagged AML1 was co-transfected into 293T cells with the indicated Flag-tagged Foxp3 mutants and immunoprecipitated with anti-Flag M2 agarose, and protein blots were probed with anti-HA (top) or anti-Flag (middle) antibodies. The expression of AML1 in the lysates was monitored by immunoblotting using an anti-HA antibody (bottom). WT, wild type. Abbreviations of mutants are as follows: 281RR: I281R and V282R; 298LV: R298L and E299V; 309LV: R309L and R310V; 316LD: H316L and G317D; 322RV: P322R and E323V; 329VH: D329V and Y330H; 329VHL: D329V, Y330H and K332L. **h**, Mouse $CD4^+$ T cells transduced with pMCsIg vectors encoding wild-type and mutant Foxp3 proteins, IRES and GFP were stained intracellularly with phycoerythrin-conjugated anti-Foxp3, and analysed by flow cytometry. Numbers in dot plots indicate mean fluorescence intensity (MFI) of Foxp3 in the rectangle gates. **i**, Gene-transduced T cells were stimulated as in Fig. 1h, and $CD4^+$ cells were analysed for intracellular IL-2 by flow cytometry. Error bars represent s.e.m.; all experiments were repeated at least three times with similar results.

To examine further whether IL-2 repression by Foxp3 is dependent on the interaction of Foxp3 with AML1, we made Foxp3 mutants with two or three amino acid substitutions in the region required for interaction with AML1 (Fig. 3f). In co-immunoprecipitation experiments using these Foxp3 mutants and wild-type AML1, mutant 329VHL, which has three amino acid changes in the region, scarcely bound to AML1, and mutant 329VH exhibited less AML1 binding than other mutants with two amino acid substitutions or than wild-type Foxp3 (Fig. 3g). After retroviral gene transduction of mutated or wild-type Foxp3 into conventional $CD4^+$ T cells, the expression levels of the Foxp3 proteins were equivalent (Fig. 3h). However, the AML1-non-binding 329VHL mutant was unable to suppress IL-2 production when transduced to conventional $CD4^+$ T cells, and the 329VH mutant with reduced AML1 binding was less suppressive than other mutants or wild-type expressers (Fig. 3i). 329VHL is not a simple loss-of-function mutant, because it retained the ability to interact with NFAT (Supplementary Fig. S6) and regulate some Foxp3-regulated genes in a DNA microarray analysis (Supplementary Fig. S7). Thus, in addition to the NFAT–Foxp3 interaction[22], the AML1–Foxp3 interaction is critical on its own for suppressing IL-2.

We next examined whether AML1 is involved in the regulation of the genes encoding T_R-cell-associated cell surface molecules, such as CD25, cytotoxic T-lymphocyte-associated antigen-4 (CTLA-4) and glucocorticoid-induced TNF-receptor-family-related protein (GITR). ChIP assays showed that AML1 bound to intron 1 of *CD25*, to the promoter of *CTLA-4*, and to the promoter and intron 1 of *GITR* (Supplementary Fig. S8a, b). ChIP assays using the same primer pairs as in Fig. S8a showed that Foxp3 bound to *CD25* intron 1 and *GITR* intron 1 (Supplementary Fig. S9), which was consistent with recent reports[22]. Retroviral transduction of AML1 in mouse primary $CD4^+$ T cells downregulated the expression of CD25, CTLA-4 and, in particular, GITR (Supplementary Fig. S8c), whereas similar transduction of wild-type Foxp3 upregulated the expression of these molecules (Fig. 4a). Notably, the 329VHL Foxp3 mutant failed to upregulate them, and the 329VH mutant was less efficient in the upregulation than other mutants or wild-type Foxp3 (Fig. 4a). Taken together, these findings suggest that AML1 and the Foxp3–AML1 complex control these genes in conventional T cells and natural T_R cells, respectively.

We, and others, have shown previously that natural T_R cells and Foxp3-transduced T cells are anergic (that is, non-proliferative) on stimulation of the TCR *in vitro*, whereas stimulation of the TCR together with high-dose IL-2 and anti-CD28 stimulation can abrogate the anergic state[4,5,23,24]. As shown in Fig. 4b, c, 329VHL-transduced T cells were slightly less anergic than T cells transduced with other mutants or wild-type Foxp3, and the addition of high-dose IL-2 and anti-CD28 antibody more easily abrogated their anergic state. Furthermore, whereas natural T_R cells and Foxp3-transduced T cells suppress the proliferation of co-cultured naive T cells on stimulation of the TCR *in vitro*[4,5,23,24], the 329VHL-transduced cells were much less suppressive than T cells transduced with other mutants (Fig. 4d). In addition, knockdown of AML1 in human $CD25^{high}CD4^+$ T cells abrogated their anergic state and attenuated their suppressive activity (Fig. 4e–g, and Supplementary Fig. S10). Taken together, these results support a model in which the formation of the Foxp3–AML1 complex controls anergy and the suppressive function of natural T_R cells.

AML1 and the Foxp3–AML1 complex may control the expression of a broad range of genes in addition to those encoding the IL-2 and T_R-cell-associated molecules described above (Supplementary Fig. S7). For example, AML1 and Foxp3 controlled IFN-γ production in a similar manner to that of IL-2: AML1 activated *IFN-γ* gene transcription by binding to its promoter, and Foxp3 suppressed AML1-induced IFN-γ production (Supplementary Information and Supplementary Fig. S11). Although the precise molecular mechanism of how AML1 and the Foxp3–AML1 complex control their

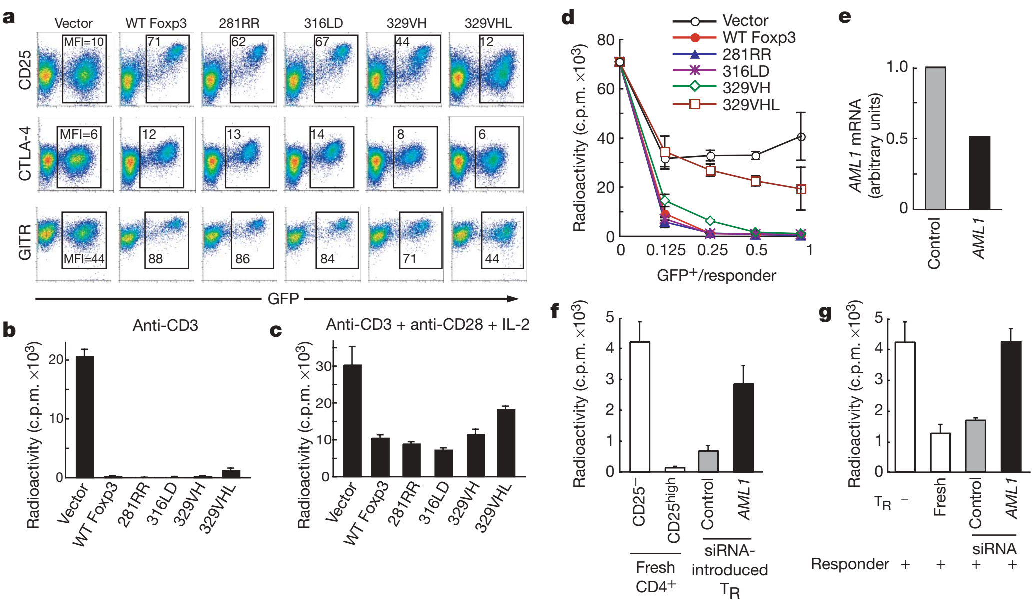


Figure 4 | Foxp3 requires interaction with AML1 to confer T_R-cell phenotype and function on conventional T cells. **a**, Mouse conventional T cells transduced with pMCsIg vector encoding wild-type (WT) or mutant Foxp3 shown in Fig. 3g–j were stained with anti-CD25 (top), anti-CTLA-4 (intracellular, middle) or anti-GITR (bottom), and analysed by flow cytometry. MFI, mean fluorescence intensity. **b**, **c**, Proliferation of sorted GFP^+ cells in the presence of soluble anti-CD3 and antigen-presenting cells (APCs) without (**b**) or with (**c**) IL-2 and anti-CD28, measured by $[^3H]$thymidine incorporation; results are means ± s.d. for triplicate cultures. **d**, Suppressive activity of T cells transduced with wild-type or mutant Foxp3. Graded doses of sorted GFP^+ cells were cultured with freshly prepared $CD25^-CD4^+$ cells for 72 h with soluble anti-CD3 and APCs, and $[^3H]$thymidine incorporation was measured as in **b**. **e**, Knockdown of AML1 results in repression of *AML1* mRNA in human primary $CD25^{high}CD4^+$ T_R cells. Relative expression of AML1 was quantified by real-time polymerase chain reaction, using HPRT as an internal control. **f**, Proliferation of control or *AML1* siRNA (shown in Fig. 1g as AML1 no. 1)-introduced human primary $CD25^{high}CD4^+$ T_R cells in the presence of soluble anti-CD3 and APCs. **g**, Suppressive activity of human primary $CD25^{high}CD4^+$ T_R cells transduced with control or *AML1* siRNA. Fresh or siRNA-transduced $CD25^{high}CD4^+$ T_R cells were mixed with freshly prepared $CD25^-CD4^+$ responder T cells, and stimulated by soluble anti-CD3 in the presence of APCs. Proliferation of cells was assessed by $[^3H]$thymidine incorporation and is shown as mean ± s.d. for triplicate cultures. All experiments were repeated at least three times with similar results.

downstream genes remains to be established, the present study provides several models for transcriptional controls by AML1 and Foxp3 (Supplementary Discussion). In addition, Foxp3 could bind to two other members of the AML/Runx protein family, AML2 (Runx3) and AML3 (Runx2) (Supplementary Discussion and Supplementary Fig. S12). There are recent reports that polymorphisms of AML1 and AML1-binding sites are associated with the susceptibility to several autoimmune diseases in humans[25–28]. Such genetic polymorphisms might affect T_R-cell-mediated maintenance of immunological self-tolerance (Supplementary Discussion). The interaction between AML1 and Foxp3 may be a potential therapeutic target for controlling physiological and pathological immune responses.

METHODS

RNA interference. Human PBMCs were freshly separated from whole blood and purified to $CD4^+$ T cells. Sorted cells were subsequently transfected with the Stealth RNAi duplex oligonucleotides against *AML1* (HSS141472; Invitrogen). Stealth RNAi negative control GC high was used as a control short interfering RNA (siRNA). At 16 h after transfection, 20 U ml^{-1} human recombinant IL-2 (eBiosciences) was added to the culture. At 40 h after transfection, $CD25^{high}CD4^+$ T_R cells (defined as the high 2% fraction of $CD4^+$ T cells) were sorted. At this time point, control and *AML1* siRNA-introduced T_R cells expressed CD25, FOXP3 and CTLA-4 at equivalent levels (Supplementary Fig. S10). $CD25^-CD4^+$ T cells were similarly sorted and used as responders. $CD3^-$ cells were irradiated (25 Gy) and used as APCs. A total of 10,000 $CD25^{high}CD4^+$ T_R cells were mixed with 20,000 APCs with or without 10,000 responder cells. Cells were stimulated with 20 ng ml^{-1} anti-CD3 (HIT3a; BD Pharmingen) for 4 days. [^{3}H]Thymidine incorporation for the last 8 h of cell culture was measured as an indicator of cell proliferation and is expressed as the mean $\pm$ s.d. for triplicate cultures.

Retroviral transduction. For co-transduction experiments, cells were first infected with pMCsIg-AML1 and kept for a further 5 h at 32 °C with 5% CO_2 and were subsequently infected with pGCSamIN–Foxp3. Gene transduction into $CD25^-CD4^+$ conventional T cells and $CD25^+CD4^+$ T_R cells was performed by stimulating cells with plate-bound anti-CD3 (10 μg ml^{-1}) and soluble anti-CD28 (1 μg ml^{-1}) with 100 U ml^{-1} mouse recombinant IL-2 for 16 h. Activated T cells were infected by resuspending cells with viral supernatants supplemented with 100 U ml^{-1} IL-2 and 5 μg ml^{-1} Polybrene, followed by centrifugation for 1 h at 3,200 r.p.m. Cells were cultured at 37 °C with 5% CO_2 for a further 24 h and sorted by Moflo (DakoCytomation) for ELISA analyses.

Received 12 December 2006; accepted 9 February 2007.
Published online 21 March 2007.

1. Shevach, E. M. Regulatory T cells in autoimmmunity. *Annu. Rev. Immunol.* **18**, 423–449 (2000).
2. Maloy, K. J. & Powrie, F. Regulatory T cells in the control of immune pathology. *Nature Immunol.* **2**, 816–822 (2001).
3. Sakaguchi, S. Naturally arising $CD4^+$ regulatory T cells for immunologic self-tolerance and negative control of immune responses. *Annu. Rev. Immunol.* **22**, 531–562 (2004).
4. Hori, S., Nomura, T. & Sakaguchi, S. Control of regulatory T cell development by the transcription factor Foxp3. *Science* **299**, 1057–1061 (2003).
5. Fontenot, J. D., Gavin, M. A. & Rudensky, A. Y. Foxp3 programs the development and function of $CD4^+$ $CD25^+$ regulatory T cells. *Nature Immunol.* **4**, 330–336 (2003).
6. Khattri, R., Cox, T., Yasayko, S. A. & Ramsdell, F. An essential role for Scurfin in $CD4^+$ $CD25^+$ T regulatory cells. *Nature Immunol.* **4**, 337–342 (2003).
7. Fontenot, J. D., Rasmussen, J. P., Gavin, M. A. & Rudensky, A. Y. A function for interleukin 2 in Foxp3-expressing regulatory T cells. *Nature Immunol.* **6**, 1142–1151 (2005).
8. Okuda, T., van Deursen, J., Hiebert, S. W., Grosveld, G. & Downing, J. R. AML1, the target of multiple chromosomal translocations in human leukemia, is essential for normal fetal liver hematopoiesis. *Cell* **84**, 321–330 (1996).
9. Wang, Q. *et al.* Disruption of the *Cbfa2* gene causes necrosis and hemorrhaging in the central nervous system and blocks definitive hematopoiesis. *Proc. Natl Acad. Sci. USA* **93**, 3444–3449 (1996).
10. Taniuchi, I. *et al.* Differential requirements for Runx proteins in CD4 repression and epigenetic silencing during T lymphocyte development. *Cell* **111**, 621–633 (2002).
11. Komine, O. *et al.* The Runx1 transcription factor inhibits the differentiation of naive $CD4^+$ T cells into the Th2 lineage by repressing *GATA3* expression. *J. Exp. Med.* **198**, 51–61 (2003).
12. Jain, J., Loh, C. & Rao, A. Transcriptional regulation of the *IL-2* gene. *Curr. Opin. Immunol.* **7**, 333–342 (1995).
13. Rooney, J. W., Sun, Y. L., Glimcher, L. H. & Hoey, T. Novel NFAT sites that mediate activation of the interleukin-2 promoter in response to T-cell receptor stimulation. *Mol. Cell. Biol.* **15**, 6299–6310 (1995).
14. Bruniquel, D. & Schwartz, R. H. Selective, stable demethylation of the interleukin-2 gene enhances transcription by an active process. *Nature Immunol.* **4**, 235–240 (2003).
15. Novak, T. J., White, P. M. & Rothenberg, E. V. Regulatory anatomy of the murine interleukin-2 gene. *Nucleic Acids Res.* **18**, 4523–4533 (1990).
16. Kitabayashi, I., Yokoyama, A., Shimizu, K. & Ohki, M. Interaction and functional cooperation of the leukemia-associated factors AML1 and p300 in myeloid cell differentiation. *EMBO J.* **17**, 2994–3004 (1998).
17. Yagi, H. *et al.* Crucial role of FOXP3 in the development and function of human $CD25^+$ $CD4^+$ regulatory T cells. *Int. Immunol.* **16**, 1643–1656 (2004).
18. Walker, M. R. *et al.* Induction of FoxP3 and acquisition of T regulatory activity by stimulated human $CD4^+$ $CD25^-$ T cells. *J. Clin. Invest.* **112**, 1437–1443 (2003).
19. Fontenot, J. D. *et al.* Regulatory T cell lineage specification by the forkhead transcription factor Foxp3. *Immunity* **22**, 329–341 (2005).
20. Ono, M., Shimizu, J., Miyachi, Y. & Sakaguchi, S. Control of autoimmune myocarditis and multiorgan inflammation by glucocorticoid-induced TNF receptor family-related protein^{high}, Foxp3-expressing $CD25^+$ and $CD25^-$ regulatory T cells. *J. Immunol.* **176**, 4748–4756 (2006).
21. Kanno, Y., Kanno, T., Sakakura, C., Bae, S. C. & Ito, Y. Cytoplasmic sequestration of the polyomavirus enhancer binding protein 2 (PEBP2)/core binding factor α (CBFα) subunit by the leukemia-related PEBP2/CBFβ-SMMHC fusion protein inhibits PEBP2/CBF-mediated transactivation. *Mol. Cell. Biol.* **18**, 4252–4261 (1998).
22. Wu, Y. *et al.* FOXP3 controls regulatory T cell function through cooperation with NFAT. *Cell* **126**, 375–387 (2006).
23. Thornton, A. M. & Shevach, E. M. $CD4^+$ $CD25^+$ immunoregulatory T cells suppress polyclonal T cell activation *in vitro* by inhibiting interleukin 2 production. *J. Exp. Med.* **188**, 287–296 (1998).
24. Takahashi, T. *et al.* Immunologic self-tolerance maintained by $CD25^+$ $CD4^+$ naturally anergic and suppressive T cells: induction of autoimmune disease by breaking their anergic/suppressive state. *Int. Immunol.* **10**, 1969–1980 (1998).
25. Brenner, O. *et al.* Loss of Runx3 function in leukocytes is associated with spontaneously developed colitis and gastric mucosal hyperplasia. *Proc. Natl Acad. Sci. USA* **101**, 16016–16021 (2004).
26. Prokunina, L. *et al.* A regulatory polymorphism in *PDCD1* is associated with susceptibility to systemic lupus erythematosus in humans. *Nature Genet.* **32**, 666–669 (2002).
27. Helms, C. *et al.* A putative RUNX1 binding site variant between *SLC9A3R1* and *NAT9* is associated with susceptibility to psoriasis. *Nature Genet.* **35**, 349–356 (2003).
28. Tokuhiro, S. *et al.* An intronic SNP in a RUNX1 binding site of *SLC22A4*, encoding an organic cation transporter, is associated with rheumatoid arthritis. *Nature Genet.* **35**, 341–348 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. Kakino, R. Ishii and M. Yoshida for technical assistance; F. Rawle for valuable comments on the manuscript; T. Kitamura for the pMCsIg retroviral vector; M. Onodera for the pGCSamIN retroviral vector; and F. Macian for the NFAT-CA construct.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to S.S. (shimon@frontier.kyoto-u.ac.jp).

LETTERS

Nuclear cytokine-activated IKKα controls prostate cancer metastasis by repressing Maspin

Jun-Li Luo[1], Wei Tan[1], Jill M. Ricono[2], Olexandr Korchynskyi[1], Ming Zhang[3], Steven L. Gonias[2], David A. Cheresh[2] & Michael Karin[1]

Inflammation enhances tumour promotion through NF-κB-dependent mechanisms[1]. NF-κB was also proposed to promote metastatogenesis through epithelial–mesenchymal transition[2]. Yet a mechanistic link between inflammation and metastasis is missing. We identified a role for IκB kinase α (IKKα), activated by receptor activator of NF-κB (RANK/TNFRSF11A), in mammary epithelial proliferation during pregnancy[3]. Owing to similarities between mammary and prostate epithelia, we examined IKKα involvement in prostate cancer and its progression. Here we show that a mutation that prevents IKKα activation slows down CaP growth and inhibits metastatogenesis in TRAMP mice, which express SV40 T antigen in the prostate epithelium[4]. Decreased metastasis correlated with elevated expression of the metastasis suppressor Maspin[5], the ablation of which restored metastatic activity. IKKα activation by RANK ligand (RANKL/TNFSF11) inhibits Maspin expression in prostate epithelial cells, whereas repression of *Maspin* transcription requires nuclear translocation of active IKKα. The amount of active nuclear IKKα in mouse and human prostate cancer correlates with metastatic progression, reduced Maspin expression and infiltration of prostate tumours with RANKL-expressing inflammatory cells. We propose that tumour-infiltrating RANKL-expressing cells lead to nuclear IKKα activation and inhibition of *Maspin* transcription, thereby promoting the metastatic phenotype.

Prostate cancer (CaP), a heterogeneous disease, progresses from prostatic intra-epithelial neoplasia to locally invasive adenocarcinoma, and then to hormone-refractory metastatic carcinoma[6]. One in six men will be diagnosed with CaP, and one in thirty-three will die of metastatic disease[6]. Early CaP confined to the prostate can be treated[7], but no effective treatments are available for metastatic disease. To understand whether signalling pathways related to NF-κB activation, which are amenable to pharmacological inhibition[8], are involved in metastatogenesis, we examined the role of IKKα, one of the two catalytic subunits of the IKK complex[9], in CaP development and metastatic progression. To this end, we used $Ikk\alpha^{AA/AA}$ mice in which the serines whose phosphorylation is required for IKKα activation are replaced with alanines[3]. Growth of the prostate and mammary epithelia depends on sex steroids and $Ikk\alpha^{AA/AA}$ females exhibit retarded mammary gland growth during pregnancy[3]; however, $Ikk\alpha^{AA/AA}$ mice show no defects in prostate development and composition (Supplementary Fig. 1). $Ikk\alpha^{AA/AA}$ mice were crossed with TRAMP mice expressing SV40 Tag from the prostate specific probasin promoter[4]. As described[4,10], single mutant TRAMP (WT/TRAMP) mice developed CaP rather early and started dying at around 22 weeks of age, but homozygosity for the $Ikk\alpha^{AA}$ allele prolonged tumour onset and delayed mortality (Fig. 1a). Nonetheless, no discernible histological differences in primary tumours (Fig. 1b, c) and lymph node metastases (not shown) were found between the genotypes on necropsy. Although no size differences of primary CaP were detected at death (which was due to primary CaP),

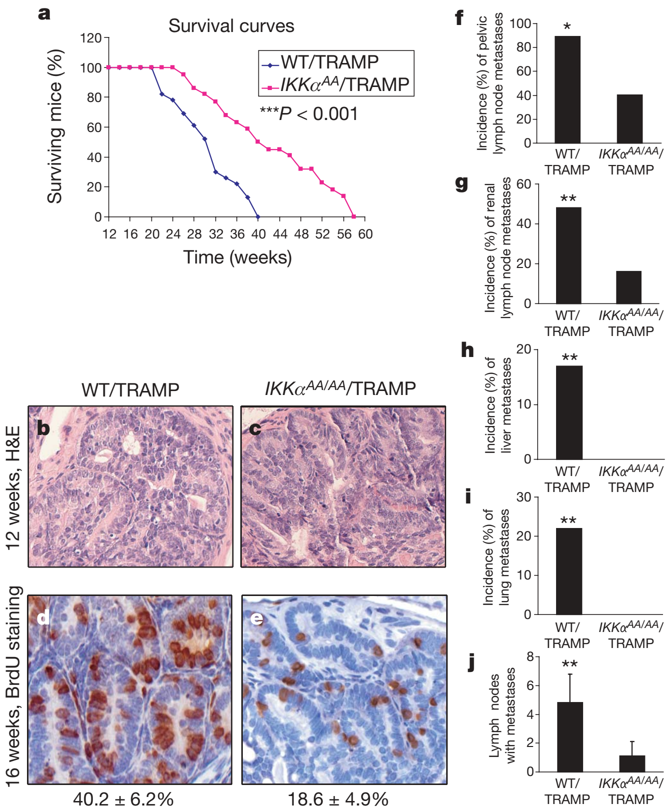


Figure 1 | IKKα activity is required for prostate cancer metastasis. **a**, TRAMP mice were intercrossed for at least six generations with $Ikk\alpha^{AA/AA}$ mice. Survival of $Ikk\alpha^{AA/AA}$/TRAMP ($n = 22$) and WT/TRAMP ($n = 23$) mice was compared. **b**, **c**, Histological analysis (haematoxylin and eosin staining of paraffin-embedded sections; magnification, ×100) of primary prostate adenocarcinomas from 3-month-old WT/TRAMP (**b**) and $Ikk\alpha^{AA/AA}$/TRAMP (**c**) mice. **d**, **e**, Cell proliferation in prostate carcinomas of 4-month-old WT/TRAMP (**c**) and $Ikk\alpha^{AA/AA}$/TRAMP (**d**) mice was examined by BrdU labelling. Percentages of BrdU-positive cells are indicated underneath ($n = 4$). **f–j**, Incidence of pelvic lymph node (**f**), renal lymph node (**g**), liver (**h**) and lung (**i**) metastases. **j**, Average numbers of lymph nodes harbouring metastases (WT/TRAMP, $n = 23$; $Ikk\alpha^{AA/AA}$/TRAMP, $n = 22$). *$P < 0.05$; **$P < 0.01$; error bars, s.d.

[1]Laboratory of Gene Regulation and Signal Transduction, Department of Pharmacology and Cancer Center, School of Medicine, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0723, USA. [2]Department of Pathology and the Moores Cancer Center, University of California, San Diego, La Jolla, California 92093-0803, USA. [3]Baylor College of Medicine, Department of Molecular and Cellular Biology, One Baylor Plaza, Houston, Texas 77030, USA.

Ikkα$^{AA/AA}$/TRAMP mice exhibited considerably fewer distant-site metastases (Fig. 1b–j).

Consistent with slow tumour growth, we observed a proliferation defect in CaP of 4-month-old *Ikkα*$^{AA/AA}$/TRAMP mice pulsed with 5-bromodeoxyuridine (BrdU; identifies cells undergoing DNA synthesis; Fig. 1d, e). These changes in cell proliferation affected tumour size only at intermediate time points (Supplementary Table 1), and at death the primary tumours had reached the same size in *Ikkα*$^{AA/AA}$/TRAMP as in WT/TRAMP mice. Progression of CaP in WT/TRAMP mice correlates with reduced E-cadherin expression, a hallmark of epithelial–mesenchymal transition, and increased expression of the neuroendocrine marker synaptophysin[10], which is also upregulated in advanced human CaP[11]. No differences in E-cadherin or synaptophysin expression were detected between the genotypes (Supplementary Fig. 2).

Tumour metastasis is governed by genetic and epigenetic factors[12]. We compared expression of approximately 40 promoters and suppressors of metastasis in primary CaP of both genotypes by quantitative reverse transcription polymerase chair reaction (qRT–PCR) (Supplementary Fig. 3). The only gene that exhibited marked and consistent differences between the genotypes was *Maspin* (*Serpinb5*), an established metastasis suppressor (Fig. 2a). *Maspin* is a gene expressed by normal mammary epithelial cells but not by mammary carcinoma, and the forced expression of Maspin inhibits invasion and motility of breast cancer cells[5]. Ectopic Maspin expression in mammary epithelium inhibits development of lobulo-alveolar structures during pregnancy[13], resulting in a phenotype similar to the mammary phenotype of *Ikkα*$^{AA/AA}$ mice[3]. Maspin overexpression from a mammary epithelial promoter inhibited metastasis of SV40 Tag-induced mammary carcinoma[14]. Importantly, there are strong inverse correlations between Maspin expression and metastatic potential in human CaP[15–17].

Although Maspin was similarly expressed in cancerous prostates of both genotypes at 3 months of age (Supplementary Fig. 4), its

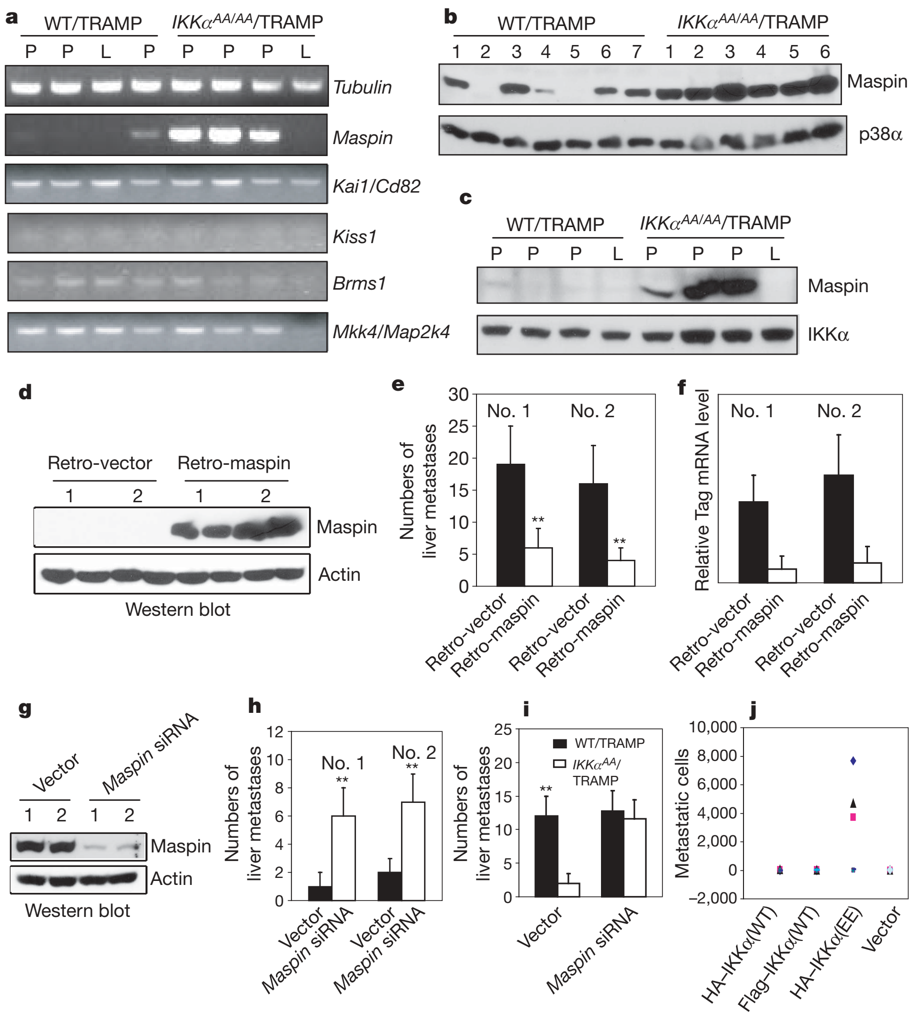


Figure 2 | Maspin expression correlates with metastatogenesis. **a**, Expression of mRNAs encoding metastasis suppressors was analysed by qRT–PCR in primary (P) CaPs and lymph node (L) metastases. **b**, **c**, Maspin expression was examined in primary CaPs of 4–5- (**b**) or 7–8- (**c**) month-old mice by immunoblotting. **d**, CaP cells from WT/TRAMP mice were infected with either a *Maspin* retrovirus (retro-maspin) or an 'empty' retrovirus (retro-vector). Two preparations of transduced cells per virus were selected in 100 μg ml^{-1} zeocin for 1 week and immunoblotted for Maspin expression. **e**, Single-cell suspensions of cells from **d** were injected into spleens of *Nu/Nu* male mice (each group, $n = 5$) and liver metastases were counted 4 weeks later. Numbers represent different cell isolates and cultures. **f**, qRT–PCR analysis of SV40 Tag mRNA in livers from **e**. **g**, Primary CaP cells from *Ikkα*$^{AA/AA}$/TRAMP mice expressing endogenous *Maspin* were infected with *Maspin* siRNA retrovirus (*Maspin* siRNA) or an 'empty' retrovirus (vector). Two preparations of transduced cells per virus were selected in 1 μg ml^{-1} puromycin and immunoblotted for Maspin expression. **h**, Single-cell suspensions of cells from **g** were injected into spleens of *Nu/Nu* male mice (each group, $n = 5$) and liver metastases were counted after four weeks. **i**, *Maspin* was knocked down in WT/TRAMP and *Ikkα*$^{AA/AA}$/TRAMP CaP cells and metastatic activity was examined as above. **j**, Four coded samples of *Ikkα*$^{AA/AA}$/TRAMP CaP cells infected with the indicated adenoviruses were inoculated on backsides of chorioallantoic membranes of 10-day-old chick embryos (each group, $n = 10$). After 9 days, pulmonary metastasis was quantified by detection of SV40 Tag DNA sequences. Chicken GAPDH DNA was used as an internal control. Error bars (**e**, **f**, **h**, **i**), s.d.

expression started declining in WT/TRAMP CaPs at 4–5 months of age but remained high in $Ikk\alpha^{AA/AA}$/TRAMP CaPs (Fig. 2b). At 7–8 months of age Maspin was not detectable in WT/TRAMP CaP, but primary $Ikk\alpha^{AA/AA}$/TRAMP CaP retained high Maspin expression (Fig. 2a, c). Immunohistochemistry revealed that Maspin was highly expressed in $Ikk\alpha^{AA/AA}$/TRAMP CaP cells, but was barely detectable or absent in WT/TRAMP CaP (Supplementary Fig. 5). Consistent with its role as a metastasis suppressor, *Maspin* expression was extinguished in the rare lymph node metastases found in $Ikk\alpha^{AA/AA}$/TRAMP mice (Fig. 2a).

To determine whether Maspin expression accounts for reduced metastasis, we examined metastatic potential of primary CaP cells from 7–8 month old mice that were cultured for 1–3 weeks. Single-cell suspensions were injected into spleens of Nude (*Nu/Nu*) male mice and 10 min later the spleens were removed. After 4 weeks, metastases to the liver were enumerated and liver SV40 Tag messenger RNA was quantified. *Maspin*-negative WT/TRAMP CaP cells gave rise to numerous liver metastases or Tag mRNA in the liver, but forced Maspin expression reduced metastatic activity (Fig. 2d–f). In contrast, $Ikk\alpha^{AA/AA}$/TRAMP CaP cells formed very few liver metastasis, but a *Maspin* knockdown (Fig. 2g) increased metastatic potential to the level of WT/TRAMP CaP cells (Fig. 2h, i). Despite marked reduction in metastatic potential, the tumorigenic potential of $Ikk\alpha^{AA/AA}$/TRAMP CaP cells was only marginally lower than that of WT/TRAMP CaP (Supplementary Fig. 6). The reduced tumour growth could be due to expression of Maspin, which can inhibit cell proliferation[16,17].

To examine whether IKKα activation modulates metastatic activity we used adenoviruses to express either wild-type (WT) or constitutively active (EE) IKKα—in which the activation loop serines were replaced with phosphomimetic glutamate residues[18]—in $Ikk\alpha^{AA/AA}$ primary CaP cells. Because adenovirus infection is transient we used the chick embryo spontaneous metastasis assay[19] to determine metastatic activity. This model examines metastatic spread of cancer cells inoculated on the chorioallantoic membrane of 10-day-old chicken embryos. Only activated-IKKα-transduced cells showed metastatic activity (Fig. 2j). Yet primary transplanted tumours grew at the same rate. Thus, in addition to Maspin, IKKα activity also determines metastatic potential.

We examined how IKKα controls *Maspin* gene expression. $Ikk\alpha^{AA/AA}$/TRAMP CaP cells expressing Maspin were infected with adenoviruses encoding green fluorescent protein (GFP), IKKα(WT), activated IKKα(EE) or activated IKKβ(EE). Three days later, Maspin levels were examined. IKKα(EE) significantly downregulated Maspin expression, whereas IKKα(WT) or IKKβ(EE) had little or no effect (Fig. 3a). To examine whether IKKα regulates *Maspin* gene transcription, we used a *Maspin*–luciferase reporter containing 759 base pairs (bp) of the human *Maspin* 5′ upstream region[20]. Transfection of the reporter with different amounts of IKKα(EE) vector into $Ikk\alpha^{-/-}$ mouse embryonic fibroblasts resulted in dose-dependent repression, an effect not seen with IKKα(EE) or IKKα(AA) (Fig. 3b). No repression of *Maspin* promoter activity was seen on co-expression of activated IKKβ(EE) (Fig. 3c), which is a more potent activator of NF-κB-dependent transcription than IKKα(EE) (Supplementary Fig. 7).

Because the transfection experiments were conducted with the human *Maspin* promoter, we used Maspin-expressing, normal, human mammary epithelial cells to examine whether ectopically introduced activated IKKα interacts with the endogenous *Maspin* promoter in native chromatin. Chromatin immunoprecipitation (ChIP) analysis revealed that activated IKKα was recruited to the endogenous *Maspin* promoter (Fig. 3d). The interaction was specific because IKKα did not interact with intron 2 of the *Maspin* gene.

Endogenous IKKα was present in both the cytoplasmic and nuclear fractions of CaP cells, whereas IKKβ and the IKKγ/NEMO regulatory subunits were exclusively cytoplasmic (Fig. 3e). Very little, if any, nuclear IKKα was found in normal prostate epithelium and immunoblotting with phospho-specific antibody revealed that nuclear IKKα in CaP was mostly in its phosphorylated and activated form. Although these results are consistent with the presence of a nuclear localization sequence (NLS) in IKKα, nuclear IKKα was previously found only in differentiating keratinocytes[21]. To examine whether IKKα acts in the nucleus to control Maspin expression, the *Maspin*–luciferase reporter was co-transfected with IKKα(EE), IKKα(EE)-NLS (a nuclear localization defective version of activated IKKα) and IKKα(KM) (a catalytically inactive mutant of IKKα) vectors into $Ikk\alpha^{-/-}$ mouse embryonic fibroblasts. Neither IKKα(EE)-NLS nor

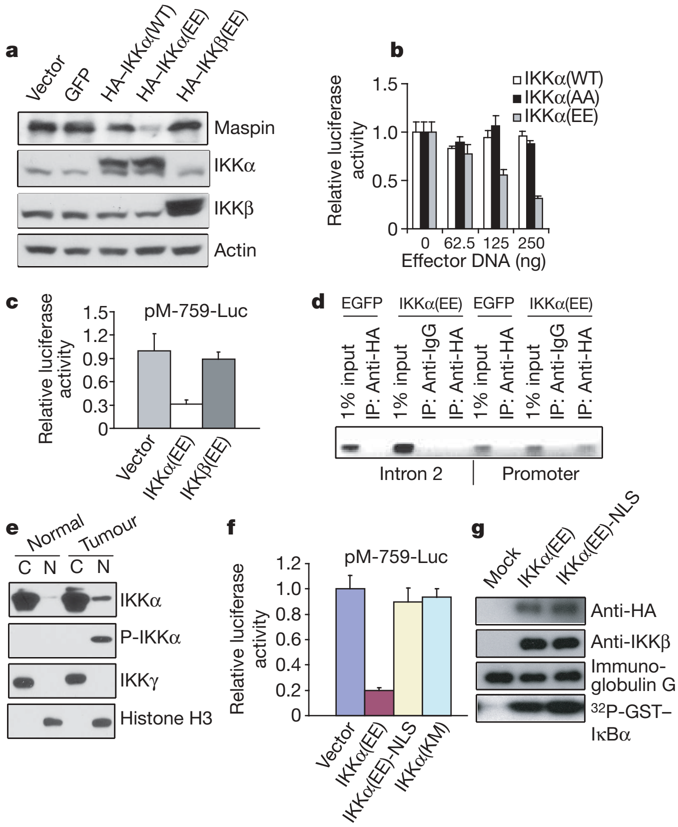


Figure 3 | Activated nuclear IKKα represses *Maspin* transcription. **a**, $Ikk\alpha^{AA/AA}$/TRAMP CaP cells were infected with different adenoviruses. After three days, Maspin, IKKα, IKKβ and actin levels were analysed by immunoblotting. **b**, A *Maspin*–luciferase reporter containing 759 bp of the human *Maspin* 5′ upstream region (pM-759-Luc) was co-transfected with different amounts of IKKα(WT), IKKα(AA) or IKKα(EE) vectors into $Ikk\alpha^{-/-}$ mouse embryonic fibroblasts. Luciferase activity was measured and normalized to a co-transfected PRL-TK reporter ($n = 3$). **c**, The *Maspin*–luciferase reporter was co-transfected with IKKα(EE) or IKKβ(EE) vectors into $Ikk\alpha^{-/-}$ mouse embryonic fibroblasts. Luciferase activity was measured and normalized as above ($n = 3$). **d**, ChIP analysis of IKKα recruitment to the human *Maspin* promoter. HME cells were transduced with either HA-IKKα(EE) or enhanced GFP (EGFP) adenovirus. After 48 h, proteins were crosslinked to DNA and chromatin was extracted and fragmented. Protein–DNA complexes were immunoprecipitated with haemagglutinin (HA) or non-specific immunoglobulin G (IgG) antibody. The presence of *Maspin* promoter region (+87 to −315) and intron 2 (+3958 to +4350) in the immunoprecipitates was examined by PCR. **e**, Nuclear (N) and cytoplasmic (C) extracts of normal and TRAMP (tumour) prostate epithelium were analysed for IKKα, phospho(activated)-IKKα (p-IKKα), IKKγ and histone H3 by immunoblotting. **f**, The *Maspin*–luciferase reporter was co-transfected with 'empty' (vector), IKKα(EE), IKKα(EE)-NLS or IKKα(KM) vectors and luciferase activity was measured as described above ($n = 3$). **g**, HEK293 cells were transfected with different expression vectors as indicated. After 36 h, IKK complexes were immunoprecipitated with HA-specific antibody and IκB kinase activity and IKKβ content were measured. Error bars (**b**, **c**, **f**), s.d. 32p-GST–IκBα, the GST (glutathione *S*-transferase) fusion protein containing the amino-terminal 50 residues of IκBα.

IKKα(KM) repressed *Maspin* promoter activity (Fig. 3f). To exclude the possibility that IKKα(EE)-NLS lost its kinase activity, IKKα(EE)-NLS or IKKα(EE) were expressed in HEK293 cells and their IκB kinase activity was measured. IKKα(EE)-NLS was as active as IKKα(EE) in IκB phosphorylation (Fig. 3g). Hence, repression of *Maspin* transcription requires nuclear IKKα kinase activity. Because NF-κB activation depends on IκB phosphorylation in the cytoplasm, and IKKβ(EE), which is a more potent activator of NF-κB than IKKα(EE), did not repress Maspin expression, repression of Maspin by IKKα seems to be NF-κB-independent.

The amount of activated nuclear IKKα correlated with CaP progression, because it was substantially higher in CaP from 7–8-month-old WT/TRAMP mice than in CaP from younger mice (Fig. 4a). Furthermore, the amount of nuclear IKKα inversely correlated with Maspin expression. Most importantly, activated nuclear IKKα correlated with progression of human CaP and was highest in stage 4 tumours, which did not express Maspin (Fig. 4b). No activated IKKα was detected in the nuclear fraction of normal human prostate or benign prostate hyperplasia, which expressed high levels of Maspin.

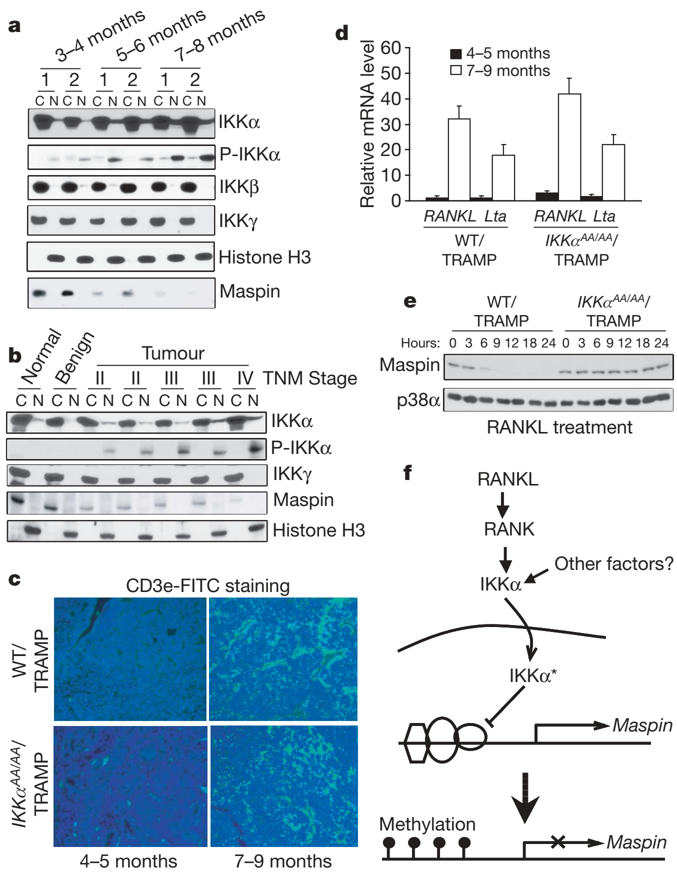


Figure 4 | Metastatic progression correlates with nuclear activation of IKKα, inflammatory cell infiltration and massive upregulation of RANKL. **a**, CaP cells from TRAMP mice of indicated ages (1 and 2 are individual mice), fractionated into cytoplasmic (C) and nuclear (N) extracts, were analysed for IKKα, phospho-IKKα, IKKβ, IKKγ, histone H3 and Maspin by immunoblotting. **b**, Normal human prostate, benign prostatic hyperplasia and prostate tumours of different clinical stages (TNM stage) were divided into nuclear (N) and cytoplasmic (C) fractions and analysed for the indicated proteins by immunoblotting. **c**, Prostate tumours from 4–5- and 7–9-month-old TRAMP mice of the indicated genotype were stained for the T cell marker CD3. FITC, fluorescein isothiocyanate. **d**, Prostate tumours from mice of the indicated age and genotype were analysed by qRT–PCR for expression of *RANKL* and *Lta* mRNAs, (results are averages ± s.d.; $n = 4$). **e**, WT/TRAMP and $Ikk\alpha^{AA/AA}$/TRAMP prostate epithelial cells were treated with lipopolysaccharide-free RANKL (200 μg ml^{-1}). At the indicated times cell extracts were prepared and examined for Maspin and p38α content by immunoblotting. **f**, A model explaining how RANK signalling leads to repression of *Maspin* transcription. After transient IKKα-mediated repression, *Maspin* transcription is likely to be permanently silenced through DNA methylation.

In mammary epithelial cells[3] and myeloid cells[22,23], IKKα is activated on occupancy of RANK by RANKL. Another cytokine that activates IKKα is lymphotoxin (LT)αβ (ref. 22). These tumour necrosis factor (TNF) family members are expressed by lymphoid and myeloid cells. Immunohistochemistry revealed only small amounts of such cells in early CaP of either genotype, but the amount of tumour infiltrating T cells ($CD3^+$) and macrophages ($F4/80^+$) was vastly increased in primary tumours of 7–9-month-old mice (Fig. 4c; Supplementary Fig. 8). Likewise, *RANKL* and *Lta* mRNA were dramatically elevated in CaP from 7–9-month-old TRAMP mice relative to tumours of younger mice (Fig. 4d).

Given the striking increase in *RANKL* mRNA, which was also paralleled at the protein level (Supplementary Fig. 9), we treated primary prostate epithelial cells with RANKL and examined Maspin expression. Maspin expression declined within 6 h of RANKL application to WT/TRAMP cells but no effect was seen in $Ikk\alpha^{AA/AA}$/TRAMP cells (Fig. 4e). These results suggest that prostate cancer metastasis is a consequence of tumour infiltration by RANKL-expressing inflammatory cells that activate IKKα in the nuclei of carcinoma cells to repress *Maspin* transcription (Fig. 4f).

In recent years it has become clear that inflammation and a pro-inflammatory microenvironment make important and critical contributions to tumour development[24]. Mechanistic studies have revealed an important tumour-promoting role for the inflammation-responsive IKK complex and its target NF-κB, acting both within cancer (or pre-malignant) cells and inflammatory cells[1]. Although inflammation is expected to enhance metastatic progression, distinct genetically established mechanisms linking inflammation and metastasis are scarce. Here we describe a novel mechanism in which IKKα activation, which can be achieved on binding of the pro-inflammatory cytokine RANKL to its receptor RANK, promotes prostate cancer metastasis (Fig. 4f). Although the $Ikk\alpha^{AA}$ mutation also affects growth rate of primary cancer, its most pronounced effect is on metastatogenesis. Analysis of genes that enhance or suppress metastasis[12] revealed that IKKα exerted its pro-metastatic effect by repressing transcription of the *Maspin* gene. Inactivation of IKKα increased Maspin expression and inhibited metastasis, whereas short interfering RNA (siRNA)-mediated *Maspin* knockdown elevated the metastatic potential of $Ikk\alpha^{AA/AA}$/TRAMP CaP cells to that of WT/TRAMP CaP cells. Repression of Maspin expression requires nuclear translocation of catalytically active IKKα. An excellent correlation between active nuclear IKKα, the level of Maspin expression and tumour progression was observed both in mouse and human CaP.

Maspin is a member of the serpin family with well-established anti-metastatic activity in breast and prostate cancers[5,17,25]. An excellent inverse correlation between Maspin expression and metastatic potential of human CaP was detected, such that metastatic CaPs express little or no Maspin[15,16]. Furthermore, in metastatic human CaP the *Maspin* promoter and 5′ control region are heavily methylated and treatment with DNA methyl transferase inhibitors reactivated *Maspin* transcription[26]. Yet, how *Maspin* transcription is repressed before epigenetic silencing occurs was unknown until now. On the basis of a general model, in which DNA methyl transferases are recruited to gene regulatory regions through specific repressors or co-repressors, the repressive action of which may be transient[27], we propose that initial repression of *Maspin* transcription is mediated by IKKα activation in response to engagement of RANK or similar receptors. With time, transient repression is converted to epigenetic silencing through DNA methylation (Fig. 4f). This process commits CaP cells to a metastatic fate.

RANKL-mediated RANK activation induces migratory behaviour in breast and prostate carcinomas and promotes bone metastasis of melanomas[28]. Our results indicate that RANK may be a general

promoter of metastatic behaviour in prostate or mammary carcinoma cells. At least in TRAMP mice, the prometastatic activity of RANK is accomplished in part through repression of *Maspin* gene transcription. Importantly, RANK is activated by RANKL, the expression of which is vastly upregulated in late stage CaP. At that time, the primary tumours are highly infiltrated with T cells and macrophages, cells that could be a major source for RANKL and related factors. Correspondingly, late-stage human and mouse prostate carcinomas contain the highest levels of active nuclear IKKα and produce little or no Maspin.

METHODS

A detailed Methods section is available in Supplementary Information. Briefly, previously described $Ikk\alpha^{AA}$ and TRAMP mice were intercrossed for six generations to generate $Ikk\alpha^{AA/AA}$/TRAMP and $Ikk\alpha^{+/+}$/TRAMP (WT/TRAMP) mice of nearly identical genetic background. Human material was obtained from the Cooperative Human Tissue Network (CHTN) along with pathology reports. Histology, gene expression and NF-κB and IKK signalling were analysed as described[3,29]. Metastatic activity was assayed either by injection of cancer cells into the spleen and measurement of liver metastases one month later[30], or by using the chick embryo metastasis assay[19].

Received 5 December 2006; accepted 2 February 2007.
Published online 18 March 2007.

1. Karin, M. Nuclear factor-κB in cancer development and progression. *Nature* **441,** 431–436 (2006).
2. Huber, M. A. *et al.* NF-κB is essential for epithelial–mesenchymal transition and metastasis in a model of breast cancer progression. *J. Clin. Invest.* **114,** 569–581 (2004).
3. Cao, Y. *et al.* IKKα provides an essential link between RANK signaling and cyclin D1 expression during mammary gland development. *Cell* **107,** 763–775 (2001).
4. Greenberg, N. M. *et al.* Prostate cancer in a transgenic mouse. *Proc. Natl Acad. Sci. USA* **92,** 3439–3443 (1995).
5. Zou, Z. *et al.* Maspin, a serpin with tumor-suppressing activity in human mammary epithelial cells. *Science* **263,** 526–529 (1994).
6. DeMarzo, A. M., Nelson, W. G., Isaacs, W. B. & Epstein, J. I. Pathological and molecular aspects of prostate cancer. *Lancet* **361,** 955–964 (2003).
7. Montironi, R., Mazzucchelli, R., Scarpelli, M., Lopez-Beltran, A. & Mikuz, G. Prostate carcinoma I: prognostic factors in radical prostatectomy specimens and pelvic lymph nodes. *BJU Int.* **97,** 485–491 (2006).
8. Karin, M., Yamamoto, Y. & Wang, Q. M. The IKK NF-κB system: a treasure trove for drug development. *Nature Rev. Drug Discov.* **3,** 17–26 (2004).
9. DiDonato, J. A., Hayakawa, M., Rothwarf, D. M., Zandi, E. & Karin, M. A cytokine-responsive IκB kinase that activates the transcription factor NF-κB. *Nature* **388,** 548–554 (1997).
10. Kaplan-Lefko, P. J. *et al.* Pathobiology of autochthonous prostate cancer in a pre-clinical transgenic mouse model. *Prostate* **55,** 219–237 (2003).
11. Chuang, C. K., Wu, T. L., Tsao, K. C. & Liao, S. K. Elevated serum chromogranin A precedes prostate-specific antigen elevation and predicts failure of androgen deprivation therapy in patients with advanced prostate cancer. *J. Formos. Med. Assoc.* **102,** 480–485 (2003).
12. Steeg, P. S. Metastasis suppressors alter the signal transduction of cancer cells. *Nature Rev. Cancer* **3,** 55–63 (2003).
13. Zhang, M. *et al.* Maspin plays an important role in mammary gland development. *Dev. Biol.* **215,** 278–287 (1999).
14. Zhang, M., Shi, Y., Magit, D., Furth, P. A. & Sager, R. Reduced mammary tumor progression in WAP-TAg/WAP-maspin bitransgenic mice. *Oncogene* **19,** 6053–6058 (2000).
15. Zou, Z. *et al.* Maspin expression profile in human prostate cancer (CaP) and *in vitro* induction of Maspin expression by androgen ablation. *Clin. Cancer Res.* **8,** 1172–1177 (2002).
16. Cher, M. L. *et al.* Maspin expression inhibits osteolysis, tumor growth, and angiogenesis in a model of prostate cancer bone metastasis. *Proc. Natl Acad. Sci. USA* **100,** 7847–7852 (2003).
17. Lockett, J., Yin, S., Li, X., Meng, Y. & Sheng, S. Tumor suppressive maspin and epithelial homeostasis. *J. Cell. Biochem.* **97,** 651–660 (2006).
18. Senftleben, U. *et al.* Activation by IKKα of a second, evolutionary conserved, NF-κB signaling pathway. *Science* **293,** 1495–1499 (2001).
19. Stupack, D. G. *et al.* Potentiation of neuroblastoma metastasis by loss of caspase-8. *Nature* **439,** 95–99 (2006).
20. Zou, Z. *et al.* p53 regulates the expression of the tumor suppressor gene maspin. *J. Biol. Chem.* **275,** 6051–6054 (2000).
21. Sil, A. K., Maeda, S., Sano, Y., Roop, D. R. & Karin, M. IκB kinase-α acts in the epidermis to control skeletal and craniofacial morphogenesis. *Nature* **428,** 660–664 (2004).
22. Bonizzi, G. *et al.* Activation of IKKα target genes depends on recognition of specific κB binding sites by RelB:p52 dimers. *EMBO J.* **23,** 4202–4210 (2004).
23. Ruocco, M. G. *et al.* IκB kinase (IKK)β, but not IKKα, is a critical mediator of osteoclast survival and is required for inflammation-induced bone loss. *J. Exp. Med.* **201,** 1677–1687 (2005).
24. Balkwill, F., Charles, K. A. & Mantovani, A. Smoldering and polarized inflammation in the initiation and promotion of malignant disease. *Cancer Cell* **7,** 211–217 (2005).
25. Chen, E. I. & Yates, J. R. Maspin and tumor metastasis. *IUBMB Life* **58,** 25–29 (2006).
26. Sato, N., Fukushima, N., Matsubayashi, H. & Goggins, M. Identification of maspin and S100P as novel hypomethylation targets in pancreatic cancer using global gene expression profiling. *Oncogene* **23,** 1531–1538 (2004).
27. Di Croce, L. *et al.* Methyltransferase recruitment and DNA hypermethylation of target promoters by an oncogenic transcription factor. *Science* **295,** 1079–1082 (2002).
28. Jones, D. H. *et al.* Regulation of cancer cell migration and bone metastasis by RANKL. *Nature* **440,** 692–696 (2006).
29. Luo, J. L., Maeda, S., Hsu, L. C., Yagita, H. & Karin, M. Inhibition of NF-κB in cancer cells converts inflammation- induced tumor growth mediated by TNFα to TRAIL-mediated tumor regression. *Cancer Cell* **6,** 297–305 (2004).
30. Morimoto-Tomita, M., Ohashi, Y., Matsubara, A., Tsuiji, M. & Irimura, T. Mouse colon carcinoma cells established for high incidence of experimental hepatic metastasis exhibit accelerated and anchorage-independent growth. *Clin. Exp. Metastasis* **22,** 513–521 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements J.-L.L. was supported by the Aventis-UICC Translational Cancer Research Fellowship, the Lopiccola Fellowship of the UCSD Cancer Center and the Life Science Research Fellowship. W.T. was supported by a postdoctoral fellowship from The Susan G. Komen Breast Cancer Foundation. Work in M.K.'s laboratory was supported by grants from the NIH, the US Army Medical Research and Materiel Command and the Prostate Cancer Foundation. M.K. is an American Cancer Society Research Professor. We thank M. Stampfer for HME cells, S. Srivastava and Z. Khalkhali-Ellis for *Maspin*–luciferase reporters and H. R. Li for assistance with the statistical analysis.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.K. (karinoffice@ucsd.edu).

CORRIGENDUM

doi:10.1038/nature07715

Nuclear cytokine-activated IKKα controls prostate cancer metastasis by repressing Maspin

Jun-Li Luo, Wei Tan, Jill M. Ricono, Olexandr Korchynskyi, Ming Zhang, Steven L. Gonias, David A. Cheresh & Michael Karin

Nature 446, 690–694 (2007)

It has come to our attention that some of the control lanes in Fig. 2a (*Kai1* and *Mkk4* control lanes) may have been inadvertently duplicated during figure assembly; the other control lanes and the experimental *Maspin* lanes are correct. We have therefore repeated the experiments and obtained the same results as those in the published figure. The primary data from this experiment can be found in Supplementary Information.

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

CORRIGENDUM

doi:10.1038/nature05791

Global warming and climate forcing by recent albedo changes on Mars

Lori K. Fenton, Paul E. Geissler & Robert M. Haberle

Nature 446, 646–649 (2007)

It has been brought to our attention that there was some ambiguity in the wording of the first boldface paragraph of this Letter regarding our prediction of the change in temperature on Mars that would have resulted from changes in albedo. As is clear from Table 1, we predict a total change of 0.65 degrees Celsius over the period from the 1970s to the 1990s, rather than 0.65 degrees Celsius per year.

Published online 5 April 2007.

naturejobs

Picture a subtropical island in southeast Asia trying to cultivate science amid its palm trees. It hopes to send its young scientists abroad to train and then lure them back to labs at home, with the ultimate aim of fostering technology that could benefit its industry. It also seeks aggressive funding growth for its top institutions.

This approach could describe Singapore, but it also applies to Taiwan — although there are some key differences between the two. Singapore's research and development (R&D) infrastructure is closely controlled from the top, by an agency called A*STAR. Taiwan's approach is less centralized, with scores of universities scrambling for new funds. The difference in each country's approach is akin to comparing classical music to jazz, with Singapore's R&D orchestra led by a disciplined conductor, and Taiwan's jazz bands tending towards more free-form improvisation.

Taiwan's bands have a lot to play for — and the stakes warrant international attention. Its programme, appropriately enough, is referred to by a cavalcade of names, including 'M-top' or 'Aim Top' (for 'Aiming for the top'), 'T-12' (a reference to the multimillion-dollar, five-year government grants going to 12 universities) and '5/500' (a reference to the number of billions of Taiwanese dollars being spent over the five years).

Whatever it is called, the funding scheme has set out its goals very clearly. It aims to bring at least one Taiwanese university into the list of the world's top 100 universities within 10 years, and into the top 50 in 15–20 years. It also plans to assume a leading position in Asia for at least 10 major disciplines or research fields in 5 years, and to be able to compete with the top 50 institutions in the same fields in 10 years.

Both Singapore and Taiwan's approaches are worth watching to see how different models of science and technology can make different, but equally beautiful music — akin to listening to both Mozart and Miles Davis.

Paul Smaglik, *Naturejobs* editor

CONTACTS
Editor: Paul Smaglik
Assistant Editor: Gene Russo

European Head Office, London
The Macmillan Building,
4 Crinan Street,
London N1 9XW, UK
Tel: +44 (0) 20 7843 4961
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

European Sales Manager:
Andy Douglas (4975)
e-mail: a.douglas@nature.com
Business Development Manager:
Amelie Pequignot (4974)
e-mail: a.pequignot@nature.com
***Nature*events:**
Claudia Paulsen Young
(+44 (0) 20 7014 4015)
e-mail: c.paulsenyoung@nature.com

France/Switzerland/Belgium:
Muriel Lestringuez (4994)
UK/Ireland/Italy/RoW:
Nils Moeller (4953)
Scandinavia/Spain/Portugal:
Evelina Rubio-Morgan (4973)
Germany/Austria/The Netherlands:
Reya Silao (4970)
Online Job Postings:
Matthew Ward (+44 (0) 20 7014 4059)

Advertising Production Manager:
Stephen Russell
To send materials use London address above.
Tel: +44 (0) 20 7843 4816
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com
***Naturejobs* web development:**
Tom Hancock
***Naturejobs* online production:**
Catherine Alexander

US Head Office, New York
75 Varick Street,
9th Floor,
New York,
NY 10013-1917
Tel: +1 800 989 7718
Fax: +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Manager: Peter Bless

Japan Head Office, Tokyo
Chiyoda Building,
2-37 Ichigayatamachi,
Shinjuku-ku,
Tokyo 162-0843
Tel: +81 3 3267 8751
Fax: +81 3 3267 8746

Asia-Pacific Sales Manager:
Ayako Watanabe
e-mail: a.watanabe@natureasia.com

YANG LIU/CORBIS

Making it big in Taiwan

This small but inventive island is putting transgenics and nanotechnology to novel uses. A pay rise might be all it needs to lure its expatriate scientists home, says **Paul Smaglik**.

Made in Taiwan: in the 1980s, that label often meant cheap plastic trinkets. In the 1990s, it referred to microchips. Now, Taiwanese scientists are working to redefine that phrase once more to mean homegrown high-tech products, researched, designed and manufactured on the tiny southeast Asian island — from transgenic papayas to telemedical devices powered by nanotechnology and material sciences.

The island, with its three main cities bordered by mountain ridges, nestled in subtropical rainforests complete with orchids and banyan trees, is abuzz with energy. There's one motor scooter for every two people and they career through narrow, curving alleys under the shadow of Taipei 101, the world's tallest building — a testament to Taiwan's aspirations.

Such aspirations are grounded in reality, because many of the country's senior faculty members and administrators trained and worked in top institutions in the United States and Europe. Some come back, mid-career, to care for ageing parents, and take a pay cut in doing so (see 'Low-tech, high-tech'). But schemes to address repatriates' salary issues are gathering steam as part of a bid to internationalize its workforce and to build better facilities in targeted areas.

One of the largest drivers toward Taiwan's goals, the Industrial Technology Research Institute (ITRI), is steering that change. ITRI boasts about 6,000 researchers, of whom roughly 950 hold PhDs. "We've created about 130 companies in the past 30 years," says ITRI's president, Johnsee Lee. But he acknowledges that many of those companies have focused on contract manufacturing, licensing out technology and outsourcing. The biggest example of this phenomenon is also one of ITRI's most prominent success stories. The Taiwan Semiconductor Company, spun off from ITRI 20 years ago, makes chips for most of the world's laptop computers, and sells them under many labels, including IBM. "We have to change from being a follower to being a differentiator, by combining technology and industrial design," says Lee.

ITRI

Chasing a killer application: ITRI president Johnsee Lee.

WiMAX for wellness

A next-generation telecommunications application called WiMAX is among ITRI's most promising platforms. The wireless communication system has bigger bandwidth and wider range than existing technologies. And combining it with other systems under development — biosensors, for instance — could create a wave of telemedicine applications. For example, using biosensors along with a WiMAX device would allow a diabetes patient to constantly monitor vital information. Already 3,800 people on the ITRI campus have signed up to test-run these devices for 'wellness enhancement', Lee says.

Like many senior Taiwanese officials, Lee trained in the West — in his case at the University of Chicago and working as a physicist at Argonne National Laboratory nearby, before returning to Taiwan in 1990. And like his colleagues, he has difficulty repatriating Taiwanese scientists who have established themselves in the United States or the European Union — although ITRI has some recruitment edges over academia, he says.

"We recruit from industry, our people leave for industry, so our compensation has to be competitive," Lee says. "We have no problem recruiting new graduates. But we have problems recruiting people with industrial experience." This year, ITRI will offer twice the going rate to key leaders and researchers. "Through that, we hope more and more experienced people will join ITRI," Lee says. ITRI also has an international intern programme to attract scientists from outside Taiwan. As another incentive, it lets its inventors receive 25% of a patent's royalties, for up to 10 years

ITRI

Grass is greener: Taiwan's Industrial Technology Research Institute hopes to tempt expat scientists back home.

after they leave. "That way it's almost like their own business," Lee says.

Academia Sinica, Taiwan's largest and most prestigious academic research institution, is facing the same problems, but with even more challenges. Its president, Chi-Huey Wong, laments the lack of infrastructure and the low pay for Taiwanese scientists. He launched his chemistry career in the United States before returning home in 2003.

"If we could just get 1–2% more back," Wong says, noting that only about 28% of Taiwanese scientists return after training abroad, compared with 98% of Japanese PhDs. "My biggest concern is salary," he adds. He is seeking sustained annual funding increases for the US$400-million institute, which he hopes will help raise pay. Although Taiwanese research institutions don't have formal schemes to offer higher salaries to Western-trained researchers or teaching faculty, flexibility is "certainly there", says Wong. For example, the president of the National Taiwan University (NTU), Si-Chen Lee, recently tendered an offer to a Nobel laureate chemist that carries an annual salary of US$210,000 — "about 2.5 times the regular salary earned by our rank-and-file professors", says Wong.

LOW-TECH, HIGH-TECH

S. C. Lee, a professor in the Institute of Applied Mechanics at the National Taiwan University (NTU), represents the current wave of Taiwanese academicians. Like most, he spent time in the United States — along with his two siblings. He came back to his homeland "because my parents are here", he says. Taiwanese scientists who return generally do so to care for ageing family members. His exposure to Silicon Valley and his roots in Taipei — where motor scooters whizz through alleys in front of street-level shops where entrepreneurs make anything from toy microscope parts to electronic components — gives him what he calls a "low-tech, high-tech" approach to physical problems.

For example, during the SARS epidemic, NTU scientists researched the best kinds of breathing masks. They examined different kinds of cloth with varying pore sizes, looked at filtration sizes necessary to keep the virus out and examined the effects of applying electrostatic charges. The 'small alley shop' ethos helped one local company create elliptical fishing gears that yield variable tension to the line, using the same technology that makes elliptical trainers such good work-out machines. The entrepreneurs are now trying to apply the physics and engineering approaches they used on the fishing reels to develop vacuum pumps. **P.S.**

Ji-Wang Chern, commissioner of research and development at the NTU, says that low pay and the relatively few academic posts help explain why few Taiwanese scientists return from abroad and most graduates go directly into industry. He aims to recruit 20 professors at US-equivalent salaries over the next five years.

Chern realizes that Taiwan can't compete with the West in broad areas such as genomics and drug discovery. But by carefully picking specialities, the country can make an impact, he says. For instance, the NTU's medical school has five scientists studying dengue virus. He aims to establish 10–15 world-class groups in similar specialities, with funding of about US$1.6 million a year per group.

Tropical transgenics

Jei-Fu Shaw, president of National Chung Hsing University (NCHU) in Taichung, shares Chern's strategic approach. "We can be number one in tropical agriculture," Shaw says. Taiwan leads the world in rice genomic research and shiitake mushroom production, and is a major exporter of tropical flowers and fruits. Transgenic varieties can offer longer shelf-life and less disease, he adds.

Shyi-Dong Yeh, vice-president of the NCHU and professor of plant pathology, is a pioneer in transgenic papayas and wants to apply that technology to aquaculture and stick to niche areas. "We don't want to compete with Monsanto," Yeh says. The NCHU, like Academia Sinica and ITRI, aims to combine technologies in focused areas, such as creating micro-sensing devices to monitor viruses in tropical crops.

To this end, the university created its Center for Nanoscience and Nanotechnology three years ago. It has already grown to 26 principal investigators, many of whom are seeking biotech applications for nano-electronic devices. Director Kuan-Jiuh Lin notes that if Taiwan can produce good microchips and flat-panel displays, the country is also capable of creating and manufacturing microelectric devices — whether for telemedicine, communications or agricultural biotech. When that happens, the country's scientists will be proud of the 'Made in Taiwan' label. ■

Paul Smaglik is editor of *Naturejobs*.

MOVERS

Johann-Dietrich Wörner, chairman of the German Aerospace Centre

1995–2007 President, Technical University Darmstadt, Germany
1990–1995 Professor of solid-structure engineering, Technical University Darmstadt, Germany
1995–2007 Professor of statics, Technical University Darmstadt, Germany

In a sense, science managers are like architects of glass buildings. "Both must aim for transparency and stability," says Johann-Dietrich Wörner, the new chairman of the German Aerospace Centre and a former building engineer.

Wörner faces considerable challenges in his new post, as the centre has lost a bit of its sparkle in recent years. Two years ago, the budget of its flagship component, the German space agency (DLR), had shrunk so much that scientists feared it would become difficult to participate in space missions at all. The agency coordinates national space projects and manages Germany's contributions to European Space Agency (ESA) missions. Pictures taken by the German-built high-resolution stereo camera on board ESA's Mars Express thrilled scientists worldwide.

The DLR has to fight to stand its ground, says Günther Hasinger, a director at the Max Planck Institute for Extraterrestrial Physics in Garching. He is optimistic that Wörner can meet the challenge. "The national space budget for the next years has happily increased, not least thanks to the community's loud complaining," he says. "I hope Wörner will treat space research with the priority it deserves."

Wörner knows it will be hard to meet all expectations. Space research and astronomy have to compete with other DLR-funded activities such as aeronautics and transport- and energy-related research. With more than 5,000 staff, the DLR is Germany's biggest research institution.

Wörner gathered management experience during the period he spent as president of the Technical University Darmstadt, one of Germany's most progressive universities in terms of academic and administrative autonomy. He earned his PhD at the university, and took a post as professor there. His presidency at Darmstadt, which began in 1995, marked his shift from passionate engineer to diplomatic science manager.

The DLR's offer came just at the right time. "I had achieved my goals at Darmstadt," he says. "It was time to move on." As president of the only German university that has a special law affording it a higher-than-usual degree of autonomy, he enjoyed considerable leeway in appointing professors and setting up new programmes.

But Wörner knows that, given its size and the scope of its mission, the DLR is in a different league. He will have to act as a mediator for ESA, the German government and the country's space-science community — a task, he says, that will require diplomacy, sensitivity and a thick skin. ■

Sophie Stigler

NETWORKS & SUPPORT

Minority report

Doing science requires a certain level of perseverance, a great deal of independence and self-motivation, and a whole lot of optimism. Coincidentally, these characteristics are epitomized by most of the immigrant workers that come to the United States in search of jobs and a better quality of life. Paradoxically, these people are very underrepresented in science.

That is why the Howard Hughes Medical Institute Gilliam Fellowship is helping underrepresented students succeed in science. It offers the scientific community a different perspective on life and research.

In June, I will be among the first group of students to graduate from the new biophysics department at the University of California, Los Angeles (UCLA), having worked on a number of research projects including fungal genomics, immunobiology, imaging and most recently cancer immunotherapy under the mentorship of Manuel Penichet. My training has been fruitful, resulting in a number of co-authored publications, a first-authored article and an American Association for Cancer Research-Thomas J. Bardos Award. My two-year project was also my inspiration for the proposal that helped me earn a Gilliam fellowship.

But research was not always part of my agenda. My goals were shaped by participation in programmes such as the National Institutes of Health-sponsored MARC (minority access to research careers) programme, which has sent minority students to graduate studies in top universities. This autumn I will begin graduate studies as the first Whitcome Fellow in the Molecular Biology Interdepartmental PhD programme at UCLA. My plan is to master the intricacies of multidisciplinary research, combining quantitative approaches and molecular biology to address important fundamental questions that lead to a better understanding of disease.

Having a Gilliam fellowship has opened many doors for me, as it has for many other motivated and independent students. I hope that other programmes will also seek out underrepresented students and further diversify the scientific community. Increasing the presence of minorities in science will provide valuable input from people with diverse points of view. Likewise, minority communities will gain from the wealth of opportunities that science careers can provide. ■

Jose Rodriguez is an undergraduate senior at UCLA.

POSTDOC JOURNAL

The benefactor

Having been educated at a women's college, it's a bit strange that my greatest ambition as a postdoctoral fellow is to be a 'kept' woman. I openly admit that one of my priorities is to satisfy my 'benefactor' — that is, the foundation that funds my external postdoctoral fellowship.

In truth, I groomed myself for this role. I worked diligently to make myself attractive to potential sponsors. The likelihood of being overlooked or rejected was high, so I learned how best to sell myself. I cultivated the qualities that would help me stand out from other appealing candidates, and I consciously made the effort to attract the right external support. After all, promises of increased financial security, additional resources, subsidized travel and a personal allowance were powerful incentives.

Of course, such generous patronage comes with certain expectations and conditions. I'm obliged to report what I do, where I travel and what I spend. I'm encouraged to always be at my personal best and to prove myself to be a good choice, a safe bet and a valuable investment. In short, I'm required to perform. Nonetheless, I'm proud of negotiating the financial arrangements that provide for my needs. I have no regrets about my ambitious and deliberate pursuit of an individual postdoctoral fellowship to support my research goals. ■

Maria Ocampo-Hafalla is a research fellow at Cancer Research UK's London Research Institute.

The inside track from academia and industry

Found in translation

For those who wish to marry research with clinical applications, finding the right environment can be hard.

David Shaywitz

Hayes Dansky

The appeal of the career path as a physician–scientist is perhaps matched only by the challenge of the career itself. Many trainees ultimately discover that the elegant concept of uniting the rigour of basic science with the compassion of patient care can be remarkably unwieldy in practice.

The need to develop — and support — translational researchers is acknowledged by leading academic medical centres, as well as by the roadmap set out by the US National Institutes of Health (NIH). "It is more and more difficult to recruit, mentor, train and retain a critical mass of clinical and translational scientists," NIH director Elias Zerhouni wrote in 2005, adding: "The increasingly complex resources needed to conduct modern clinical and translational research are either missing or scattered."

Zerhouni's description rings true. In our academic medical centres, those primarily interested in basic research or clinical care can readily find exceptional role models and well-defined career paths. Yet the aspiring translational researcher often feels cast adrift, wondering when the long-anticipated integration of science and medicine is finally going to happen.

As much as we enjoyed the basic science with which we were engaged, our determination to pursue translational research led both of us, very early in our careers, to leave attractive, well-funded academic positions to join the newly formed Department of Experimental Medicine at Merck Research Labs. On balance, our department has grown less by a formal recruiting process, and more by the self-selection of like-minded physician–scientists and colleagues who share a passion for turning promising science into clinical application.

In our experimental medicine division, we design early-stage clinical trials to evaluate emerging compounds developed against novel targets or involving new mechanisms. Our goal is to understand the fundamental science as well as the condition to be treated, so that we can determine a way to evaluate whether the molecule is working as designed and having the anticipated effect.

We work extensively with biomarkers and a range of imaging techniques. We seek to develop clinical platforms and diagnostic technologies that will allow us to examine the biological activity and safety of novel compounds when they are introduced into people. These new translational models provide a window into human physiology, generating fresh scientific insight and stimulating further experimental questions.

Above all, it is gratifying to participate directly in the development of promising new treatments. But we were surprised to discover that our ability to recognize that a compound is not working can be equally important. Across the drug industry, about 90% of compounds that enter clinical testing will fail; about a third of this attrition reflects a lack of efficacy. So the ability to make this determination early, and to divert resources into the design of better compounds rather than into additional clinical investigation of a flawed molecule, will ultimately increase the chances of bringing forward a truly useful drug.

"Fundamentally, our work is team-based, and collaboration is absolutely required for success."

Much of the excitement and satisfaction in our division comes from working with colleagues in many different regions of the organization, involved in every operational area, from basic discovery and medicinal chemistry through to manufacturing and marketing. Fundamentally, our work is team-based, and collaboration is absolutely required for success.

Of particular value to us are our relationships with colleagues in academia. Often, these are long-standing connections; but through our work in the experimental medicine division, we are increasingly meeting, discussing science and establishing collaborations with many researchers from around the world. These interactions broaden our perspective, emphasize the complexity of our shared challenges and highlight the importance of collaborative approaches.

Our fondest ambition is that the foundation we've developed will help catalyse the evolution of translational research into an integrated scientific discipline. We have established an infrastructure that can provide an expansive range of profiling and imaging platforms to our clinical investigators — capabilities that tend not to be present in most academic research environments. We are eager to contribute both our experience and our technology, and look forward to developing relationships with university investigators who share our passion for translational science.

Leaving academia was a difficult decision for both of us. But it is comforting to feel that in experimental medicine we have found an intellectual sanctuary where our vision of translational research is valued, our enthusiasm shared, our skills developed and our aspirations nurtured. After years of searching, it's nice to finally feel at home. ■

David Shaywitz and Hayes Dansky are associate directors, Department of Experimental Medicine, Merck Research Labs, Rahway, New Jersey, USA.